

<https://helda.helsinki.fi>

Highly variable species distribution models in a subarctic stream metacommunity : Patterns, mechanisms and implications

de Mendoza, Guillermo

2018-01

de Mendoza , G , Kaivosoja , R , Grönroos , M , Hjort , J , Ilmonen , J , Kärnä , O-M ,
Paasivirta , L , Tokola , L & Heino , J 2018 , ' Highly variable species distribution models in a
subarctic stream metacommunity : Patterns, mechanisms and implications ' , Freshwater
Biology , vol. 63 , no. 1 , pp. 33-47 . <https://doi.org/10.1111/fwb.12993>

<http://hdl.handle.net/10138/309490>

<https://doi.org/10.1111/fwb.12993>

unspecified

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

**Highly variable species distribution models in a subarctic stream
metacommunity: patterns, mechanisms and implications**

Guillermo de Mendoza^{1,2,*}, Riikka Kaivosoja³, Mira Grönroos⁴, Jan Hjort³, Jari Ilmonen⁵,
Olli-Matti Kärnä³, Lauri Paasivirta⁶, Laura Tokola³ & Jani Heino⁷

¹Centre for Advanced Studies of Blanes, Spanish National Research Council (CEAB-CSIC),
Accés a la Cala St. Francesc 14, ES-17300 Blanes, Spain.

²Laboratoire GEODE, UMR 5602 CNRS, Université Toulouse-Jean Jaurès, 5 allées Antonio
Machado, FR-31058 Toulouse, France.

³University of Oulu, Geography Research Unit, P.O. Box 3000, FI-90014 Oulu, Finland.

⁴University of Helsinki, Department of Environmental Sciences, Section of Environmental
Ecology, Niemenkatu 73, FI-15140 Lahti, Finland.

⁵Metsähallitus, Natural Heritage Services, P.O. Box 94, FI-01301 Vantaa, Finland.

⁶Ruuhikoskenkatu 17, FI-24240 Salo, Finland.

⁷Finnish Environment Institute, Natural Environment Centre, Biodiversity, Paavo Havaksen
Tie 3, FI-90570 Oulu, Finland.

* Corresponding author: guillermo.de-mendoza@univ-tlse2.fr

Summary

1. Metacommunity theory focuses on assembly patterns in ecological communities, originally exemplified through four different, yet non-exclusive, perspectives: patch dynamics, species sorting, source-sink dynamics, and neutral theory. More recently, three exclusive components have been proposed to describe a different metacommunity framework: habitat heterogeneity, species equivalence, and dispersal. Here, we aim at evaluating the insect metacommunity of a subarctic stream network under these two different frameworks.
2. We first modelled the presence/absence of 47 stream insects in northernmost Finland using binomial generalised linear models (GLMs). The deviance explained by pure local environmental (E), spatial (S), and climatic variables (C) was then analysed across species using beta regression. In this comparative analysis, site occupancy, as well as taxonomic and biological trait vectors obtained from principal coordinate analysis, were used as predictor variables.
3. Single-species distributions were better explained by in-stream environmental and spatial factors than by climatic forcing, but in a highly variable fashion. This variability was difficult to relate to the taxonomic relatedness among species or their biological trait similarity. Site occupancy, however, was related to model performance of the binomial GLMs based on spatial effects: as populations are likely to be better connected for common species due to their near ubiquity, spatial factors may also explain better their distributions.
4. According to the classical four-perspective framework, the observation of both environmental and spatial effects suggests a role for either mass effects or species sorting constrained by dispersal limitation, or both. Taxonomic and biological traits, including the different dispersal capability of species, were scarcely important, which undermines

the patch dynamics perspective, based on differences in dispersal ability between species. The highly variable performance of models makes the reliance on an entirely neutral framework unrealistic as well. According to the three-component framework, our results suggest that the stream insect metacommunity is shaped by the effect of habitat heterogeneity (supporting both species-sorting and mass effects), rather than species equivalence or dispersal limitation.

5. While the relative importance of the source-sink dynamics perspective or the species-sorting paradigm cannot be deciphered with the data at our disposal, we can conclude that habitat heterogeneity is an important driver shaping species distributions and insect assemblages in subarctic stream metacommunities. These results exemplify that the use of the three-component metacommunity framework may be more useful than the classical four perspective paradigm in analysing metacommunities. Our findings also provide support for conservation strategies based on the preservation of heterogeneous habitats in a metacommunity context.

Key-words

beta regression, comparative analysis, insects, metacommunity theory, single-species distribution models, stream macroinvertebrates, subarctic streams.

Introduction

Metacommunity theory predicts the assembly of ecological communities according to different perspectives. Originally, this idea was illustrated by Leibold et al. (2004) in the form of four metacommunity perspectives: (1) patch dynamics, which is based on a resource competition-colonisation trade-off among species, thus taking into account species' dispersal potential (Hanski, 1994); (2) species-sorting along environmental gradients, which relies on differences in environmental tolerance among species (Leibold, 1995); (3) mass effects or source-sink dynamics, whereby species may survive in poor-quality habitats owing to constant immigration from the source populations in high quality habitats (Pulliam 1988); and (4) the neutral theory, where demographic stochasticity solely explains assembly patterns (Hubbell, 2001). Deciphering which of these perspectives is more suitable in the context of metacommunity analysis seems difficult and may well depend on the context of analysis (e.g. spatial extent, biogeographic region, ecosystem type and more; Heino et al., 2015).

Nevertheless, the examples of metacommunity perspectives depicted in Leibold et al. (2004) are not mutually exclusive, and represent a fraction of possibilities which can be expanded with the inclusion of species dispersal rates, connectivity, species interactions, disturbance, priority effects, rapid local adaptation, meta-ecosystem dynamics and more (Brown, Sokol, Skelton, & Tornwall, 2017; Logue, Mouquet, Peter, & Hillebrand, 2011). The more recent proposal by Logue et al. (2011) claims that the metacommunity concept is better generalised by three major exclusive components, which decompose the metacommunity framework into (1) environmental heterogeneity, whereby habitat patches differ in environmental attributes; (2) species equivalence, in terms of niche characteristics; and (3) dispersal, referred to as the rate of dispersal among patches. Here, we aim at evaluating species distributions in a subarctic stream insect metacommunity under these two different

frameworks (i.e., Leibold et al., 2004 versus Logue et al., 2011), specifically so as to evaluate which of the two is more adequate for the interpretation of our observations.

Species distribution models have previously been used to predict community-level properties such as biodiversity (Ferrier & Guisan, 2006). Their accuracy in predicting community-level properties appears to be higher than that of community assembly models, although at a high cost in terms of model complexity (Bonthoux, Baselga, & Balent, 2013, Chapman & Purse, 2011). The accuracy of single-species distribution modelling, however, may also be advantageous to test ecological theories about community assembly mechanisms. This is because accurately modelling the distribution of single species, one at a time, provides the opportunity to proceed with a subsequent comparative analysis across species. Using a comparative analysis, the variation in model performance can be related, for example, to species traits and potential phylogenetic constraints.

Stream insect species, in particular, are highly suitable to decipher community assembly processes through the comparative analysis of single-species distribution models (Heino & de Mendoza, 2016). This is because of the high variability among species in tolerance of environmental conditions, as well as resource exploitation, dispersal capability, and habit traits (Merritt & Cummins, 1996; Tachet, Richoux, Bournaud, & Usseglio-Polatera, 2010; Schmidt-Kloiber & Hering, 2015; Serra, Cobo, Graça, Dolédec, & Feio, 2016). This variability is valuable in evaluating which community assembly mechanism dominates in each particular context of analysis. Basically, such an analysis might shed light into the relevance of environmental variables, spatial variables, and dispersal capability of species on model performance. Subsequently, this information can be used as an indicator of the preponderance of one community assembly mechanism over another (Figure 1). For example, if many species show similar spatial patterns, and if these species share the same dispersal potential, we can presume that the ability to disperse may be underlying the observed general

pattern for these species. This would give us hints about the adequacy to consider one particular metacommunity theory perspective over the others. Within the classical four-perspective framework (Leibold et al., 2004), patch dynamics would likely be suitable in this case, as this perspective relies on the different capability of species to both disperse and exploit resources. Within the metacommunity framework based on three exclusive components (Logue et al., 2011), dispersal would be main driver in this case. Moreover, stream insects are also a diverse group of species, which belong to different insect orders and vary widely in physiological and morphological adaptations (Merritt & Cummins, 1996). Thus, modelling the distribution of single stream insect species and subsequently proceeding with a comparative analysis across species is also a suitable indirect practice to explore possible evolutionary constraints on community assembly processes.

In this study, we analysed the distribution of common stream insect species in the metacommunity of a subarctic drainage basin. Species differ widely in their dispersal capability (e.g. passive or active dispersers, aquatic or aerial adults) and tolerance of environmental conditions such as temperature, water flow, or habitat characteristics (Grönroos et al., 2013; Heino, 2005; Heino & Grönroos, 2014). We used environmental, climatic and spatial variables as predictors of the distributions of single stream insect species. Our aim was to elucidate, first, whether or not environmental and spatial factors are relevant for explaining the distribution of stream insect species; and second, whether or not the obtained models can be related to the different dispersal capability, site occupancy (i.e. a gradient of rarity-commonness), and biological and taxonomic traits, of stream insect species. Both considerations were used to evaluate which of the two different metacommunity frameworks, either the one based on four non-exclusive perspectives (Leibold et al., 2004) or the one based on three exclusive axes (Logue et al., 2011), is more adequate to interpret our observations of single species distributions in stream networks (Figure 1).

Methods

Study area

The field work for this study (Fig. S1) was conducted in the Tenojoki drainage basin (main stem length: 361 km, basin area: 16377 km², altitude of sites: from 19 to 285 m a.s.l.) in northernmost Finland (70°N, 27°E). This subarctic drainage basin is close to a natural state, since it is characterised by very small human populations and subsequent little impact from human development. A typical feature of the area are short cool summers and long cold winters (from early November to end of May). The mean annual temperature is about -2°C in the continental areas of the drainage basin, and close to 0°C near the Arctic Ocean (Dankers & Christensen, 2005). Annual precipitation ranges from 310 mm to 410 mm depending on the location in the drainage basin (Mansikkaniemi, 1970). Most of the rainfall and snowmelt enters streams and rivers, as evaporation is generally of minor importance. Vegetation is dominated by mountain birch (*Betula pubescens* ssp. *czerepanowii*) woodlands at low altitude and barren tundra at higher altitude, but also peatlands, heathlands and riparian meadows occur commonly. Coniferous pine (*Pinus sylvestris*) woodlands occur only in scattered locations, mostly in the southern parts of the drainage basin. Wadeable streams and rivers (i.e. channel width < 25 m, water depth < 50 cm) in the area are close to a pristine state, providing excellent possibilities for examining species distributions in natural environmental conditions. We sampled altogether 55 tributary streams for this study (for details, see Kärnä et al., 2015). All these 1st to 5th order tributaries drain into the mainstem of the River Tenojoki or the River Utsjoki, and no site is located in the two mainstem rivers (Fig. S1).

Field sampling of stream insects

We took a 3-minute kick-net sample (net mesh size: 0.3 mm) at each study site (Kärnä et al., 2015) at the same time with the environmental measurements in early and middle of June 2012 (see below). The sample for each site consisted of six 30-s subsamples that were divided between main habitats at a riffle site (*ca.* 50 m²) based on visual inspections of variation in depth, flow, moss cover and particle size. The six subsamples were pooled in the field to obtain a composite sample. Such a sampling method has been shown to be effective in northern streams, allowing to detect patterns in community structure (Heino, Ilmonen, & Paasivirta, 2014) and distributions of single species (Heino & de Mendoza, 2016). Samples were immediately preserved in ethanol in the field and were taken to the laboratory for further processing and identification. Animals were separated from detritus and moss fragments and identified to the lowest possible taxonomic level, mostly species (Kärnä et al., 2015).

Species considered and species traits

We detected 107 insect taxa, of which 87 could be taxonomically determined to species or species group (Kärnä et al., 2015). Insects determined to genus level were discarded as they were considered too likely to include a few species, which is inappropriate to model single-species distributions. Then, we focused on 48 species that occurred at more than 10% of the 55 study sites, that is, that occurred in at least six sites. This is because modelling the distribution of species present in less than six sites is likely to produce spurious results and therefore the analysis of these species was considered unreliable (e.g. Pearce & Ferrier, 2000). In practice, we could model the occupancies of only 47 species because the mayfly *Baetis rhodani* occurred at all sites, so we could not use this species to model presence/absence. The 47 stream insect species considered in this study are listed in Table S1.

Nomenclature generally follows de Jong et al. (2014) and more specific references for the Simuliidae (Adler & Crosskey, 2016 ; Ilmonen, 2014).

Body size class, dispersal potential, functional feeding groups and habit trait groups were considered as species traits (Table S2). Functional feeding groups refer to exploitation of different resources, while habit traits define modes of locomotion and attachment to substrate (Merritt & Cummins, 1996). Body size classes and female dispersal potential followed a previous study (Heino & de Mendoza, 2016), with additional information from Tachet et al. (2010), Schmidt-Kloiber & Hering (2015) and Serra et al. (2016). Female dispersal potential was characterised as being “low” or “high”. In general, all species of the Simuliidae were considered to have high dispersal potential, owing to the fact that their females feed as flying adults, in most cases searching for blood of vertebrates, and hence were assumed here to generally persist much longer as active flyers than the rest of species. In this regard, Baldwin, West, and Gomery (1975) often found their marked Simuliidae females several kilometers away from their natal streams. Owing to their small size, the Simuliidae may also be distributed long distances passively by wind (Crosskey, 1990). All other species were considered as weak dispersers except for the caddisflies *Plectrocnemia conspersa* and *Potamophylax cingulatus*, according to the information available for these taxa from previous studies (Gíslason, Hannesdóttir, Munoz, & Pálsson, 2015; Hoffsten, 2004; Müller-Peddinghaus, 2011; Müller-Peddinghaus & Hering, 2013; Schmidt-Kloiber & Hering, 2015). Although such information about dispersal abilities of stream insects is rather simple, there is currently no better information available (Schmidt-Kloiber & Hering, 2015; Serra et al., 2016; Tachet et al., 2010). Functional feeding and habit trait groups generally follow Merritt & Cummins (1996).

Local environmental, climatic, and spatial variables

211 The 55 streams were surveyed during the early northern summer, between early and middle
212 of June in 2012. We measured a set of local (i.e. proximal) environmental variables that have
213 been found important for stream insects in northern drainage basins in previous studies
214 (Heino et al., 2014; Kärnä et al., 2015). These comprised physical habitat and water physico-
215 chemical variables. For physical habitat variables, we measured current velocity (m/s) and
216 depth (cm) at 30 random spots in a riffle site. We also measured mean width of the riffle site
217 based on five cross-channel measurements, evenly spaced across the surveyed riffle site.
218 Bank height and bank slope were measured at the same locations with stream width
219 measurements. Bank height was measured as the height of the lower stream bank, i.e. the
220 height from the water level to the edge of terrestrial vegetation. Bank slope was measured
221 (perpendicular to the stream) as a stream bank rise (cm) over 2 m starting from the edge of
222 terrestrial vegetation. Moss cover (%) and particle size classes (%) were visually estimated at
223 10 squares (1 m²) at random locations in a riffle site. We used a modified Wentworth's
224 (1922) scale of particle size classes: sand (0.25–2 mm), gravel (2–16 mm), pebble (16–64
225 mm), cobble (64–256 mm) and boulder (256–1,024 mm). Based on the visual estimates for
226 each square, we calculated mean values for each particle size class and moss cover at a site
227 and used these mean values in species distribution modelling. We also visually estimated
228 shading (%) by riparian vegetation and proportion of riparian deciduous trees (%). For
229 physico-chemical properties, we measured pH, conductivity and water temperature at each
230 site in the field using a YSI device model 556 MPS (YSI Inc., Ohio, USA) and took
231 additional water samples during the field campaign for further analysis. Water samples were
232 frozen at the end of the day at the Kevo Field Station situated in the northern part of the study
233 area, and were later analysed for total nitrogen, colour, iron and manganese in the laboratory
234 of the Finnish Environment Institute in Oulu following Finnish national standards (National
235 Board of Waters, 1981).

We also included three climatic variables, including annual air temperature sum above 5°C (growing degree days), mean annual air temperature and mean July air temperature for the period 1981–2010. These variables were calculated in ArcMap 10.2 for each site from a gridded (1 x 1 km) climate data provided by the Finnish Meteorological Institute (Pirinen et al., 2012). The gridded climate data were produced using meteorological station observations and Kriging interpolation (e.g. Aalto, Pirinen, Heikkinen, & Venäläinen, 2013). The selected climatic variables are likely to be important for the distributions of insects in this subarctic area, where temperature is closely associated with insect life cycles (Danks, 2007).

Spatial variables were distance-based Moran's Eigenvector Maps (db-MEM) based on geographical distances among sites (Dray, Legendre & Peres-Neto, 2006). These spatial db-MEM variables were obtained with the function "PCNM" of the R package "PCNM" (Legendre, Borcard, Blanchet, & Dray, 2013; R Core Team, 2013). We used the largest distance in the minimum spanning tree, keeping all sites connected, as the truncation threshold. Spatial db-MEM variables represent structures of autocorrelation at all spatial scales. Only those spatial db-MEM variables showing significant positive autocorrelation were included in subsequent modelling (Borcard, Gillet & Legendre, 2011), resulting in 13 spatial variables (Figure 2). Based on eigenvalues and bubble plot maps, the spatial variables can be divided into those ranging from large-scale spatial structures (e.g. V1, V2) and those showing very small scale spatial patterns (e.g. V12, V13).

Prior to modelling species distribution, we eliminated strongly correlated (i.e. Pearson $r > .7$) predictor variables from the sets of local environmental and climatic variables (see Dormann et al., 2013). Hence, we removed one variable (i.e. annual temperature sum) from the climatic variables and four variables (i.e. water iron, colour, conductivity and boulders) from the stream environmental variables. The spatial variables were already not mutually correlated (Borcard et al., 2011).

Modelling species distributions

The distribution (i.e. presence/absence) of each species was modelled using binomial generalised linear models (i.e. binomial GLMs with logit link function), using separately local environmental, climatic and spatial variables, with the R package “Rcmdr” (Fox, 2005). The deviance explained for each species was thus obtained for each binomial GLM with each of these three different subsets of variables (Figure 2). The variables selected for each species’ model were based on forward selection and Bayesian Information Criterion (BIC), separately for each variable group (i.e. environmental, climate and spatial). BIC values were used because they prevented the selection of too complex models in our case, in contrast to AIC (results not shown), which is often the case under large sample sizes (Burnham & Anderson, 2004). Moreover, the target model under BIC selection does not depend on sample size, in contrast to AIC (Burnham & Anderson, 2004). Therefore, AIC may be problematic in our case as we aim at comparing model performance between species, which may differ in the number of presences and absences. Also, deviating observations were removed from some species’ models if they had Cook’s distance values > 1 and hence affected profoundly a few models (Cook, 1977). For environmental variables, we registered whether the effect was positive or negative on species distributions. We then used the selected variables of these three subsets (i.e. local environmental, climatic and spatial) to perform variation (deviance) partitioning by subtraction, similarly as performed in multivariate contexts (Legendre & Legendre, 2012). Specifically, the deviance accounted for subset A, subset B, and subset A and B together, was computed, so as to obtain the different fractions of variation solely explained by each subset (i.e. unshared with other subsets). We eventually obtained adjusted D^2 values (Guisan & Zimmermann, 2000; Legendre & Legendre, 2012) which could be attributed to pure local environmental (E), climatic (C) or spatial effects (S), as well as to

total effects combining the three subsets of pure effects and their joint effects (E+C+S effects; Figure 2). Modelling methods other than GLMs could have been possible, yet species probably show linear responses to the environmental predictors due to the fact that they are on the edge of their geographical and ecological distributions, making GLMs adequate. Adding quadratic terms to binomial models is unlikely to change results substantially in these situations (e.g. Pulido, Riera, Ballesteros, Chappius, & Gacia, 2015), and increase the difficulty of interpretation of the results. Also, deviance partitioning is easy to accomplish when this is based on GLMs.

Comparative analysis across species

We performed a comparative analysis across species using beta regression (Ferrari & Cribari-Neto, 2004), where the adjusted D^2 values obtained with previous binomial GLMs were used as the dependent variable to be explained by site occupancy, taxonomic vectors or species trait vectors (Figure 2). These vectors were obtained separately from Principal Coordinate Analysis (PCO). Using the taxonomic relatedness of species, a taxonomic relatedness matrix was built using the function “taxa2dist” in the R package “vegan” (Oksanen et al., 2013), and taxonomic vectors were handled as continuous PCO vectors with the function “pco” in the R package “ecodist” (Goslee & Urban, 2007). The first four taxonomic eigenvectors were selected as these had much higher eigenvalues than the rest (Fig. S2). Similarly, species trait vectors were also computed using body size class, dispersal potential, functional feeding groups, and habit trait groups (Table S2). Species traits were considered as regular factors, except body size class which was considered as an ordered factor, to obtain a distance matrix based on Gower’s metric with the function “daisy” of the R package “cluster” (Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2013), and eventually trait PCO vectors with the function “cmdscale”. The four trait eigenvectors obtained were considered for further

statistical analyses. The variation in adjusted D^2 values across species that could be attributed to pure E, pure C, pure S, or E+C+S effects was fitted on site occupancy, the four taxonomic and four species trait vectors selected, using beta regression with the function “betareg” of the R package “betareg” (Cribari-Neto & Zeileis, 2010). Beta regression is adequate when the response variable (in this case, the adjusted D^2 values) is constrained between 0 and 1.

We compared the explained variation by pure E, C, and S effects with a Kruskal-Wallis test, with additional Mann-Whitney tests for subsequent pair-wise comparisons between groups. Non-parametric tests were chosen since adjusted D^2 values data departed from normality following the Shapiro-Wilk test (Zar, 1984). We also analysed the univariate relationships between site occupancy, body size, dispersal potential, broad taxonomic insect groups, functional feeding groups, habit trait groups, and taxonomic and trait vectors. Depending on the continuous (e.g. site occupancy) or categorical (e.g. habit trait group) nature of the variables involved, we followed Kruskal-Wallis tests, Mann-Whitney tests, Fisher’s exact test or Spearman correlations, as these variables were generally not normally distributed (Zar, 1984).

Results

Single species models

Local environmental and spatial effects accounted for a higher variation in species distributions (16.1% and 12.6% in average, respectively) than did climatic effects (5.4%; $p < .001$, Kruskal-Wallis test), whereas the average deviance explained did not differ significantly between local environmental and spatial effects ($p = .125$, Mann-Whitney test) (Table S3). The local environmental factors most frequently selected in explaining species distributions were water temperature, shading, and to a lesser extent, stream width, cobbles and moss (Figure 3). The spatial variables most often selected were better represented by

large-scale spatial variables within the Tenojoki drainage basin (e.g. V1, V2) than by small-spatial scale variables (e.g. V12, V13), as also shown in Figure 3. Amongst the climate variables, mean annual temperature was significant in explaining the distribution of 32 species, and July air temperature of 20 species (not shown in Fig. 3).

The adjusted deviance explained by binomial GLMs was highly variable across species and difficult to relate to particular taxonomic groups (Table S3). For example, local environmental effects were particularly relevant for the stonefly *Siphonoperla burmeisteri* (i.e. accounting for 66.7% of adjusted D^2 values), the mayfly *Heptagenia dalecarlica* (50.2%), and the blackfly *Prosimulium hirtipes* (37.7%), whereas spatial effects were most relevant for the caddisfly *Rhyacophila nubila* (41.2%), the stonefly *Brachyptera risi* (31.5%), and the chironomid midge *Cardiocladius capucinus* (28.6%). Climate effects were also highly variable. They were generally low (see above), and accounted for more than 20% of adjusted D^2 values in only three cases: the stoneflies *Diura nanseni* and *Siphonoperla burmeisteri* (34.4% and 20.1%, respectively), and the chironomid midge *Orthocladius rivicola* (28.9%). Combining all effects, binomial GLMs explained on average 37.8% of the null deviance (Table S3).

Comparative analysis across species models

The highly variable species-local environment and species-climate relationships in binomial GLMs were not accounted for by site occupancy, or by taxonomic and trait vectors, in the beta regression analysis (Table 1). The deviance explained by spatial variables was, however, significantly (i.e. $p < .05$) accounted for by site occupancy (Table 1). The influence of TAX-PCO4 and TRA-PCO2 on the adjusted D^2 values predicted by spatial effects in binomial GLMs was significant as well. Also, the influence of TAX-PCO3 was marginally significant (i.e. $p < .10$), remaining like this in the binomial GLMs based on all variables combined

(Table 1). However, when repeating the beta regression analysis by using only the significant variables selected (i.e. site occupancy, TAX-PCO3, TAX-PCO4, and TRA-PCO2), only site occupancy was statistically significant ($p = .017$), but not TAX-PCO3, TAX-PCO4 or TRA-PCO2 ($p = .943$, $p = .175$, and $p = .449$, respectively, results not shown in Table 1). Analysing through beta regression the univariate relationship of these variables with the adjusted D^2 values of binomial GLMs based on spatial effects produced a similar result (site occupancy, $p = .036$, Fig. S3; TAX-PCO3, TAX-PCO4, and TRA-PCO2, $p = .760$, $p = .660$, and $p = .524$, respectively, results not shown). This univariate relationship between the adjusted D^2 values and site occupancy was not observed when the adjusted D^2 values of binomial GLMs were referred to environmental or climate effects (Fig. S3). No statistical significance was observed either for univariate relationships between separate species traits and the adjusted D^2 values in binomial GLMs, with the sole exception of body size (Fig. S3).

The TAX-PCO3 vector showed the highest species scores for blackflies (Simuliidae) and the lowest for mayflies (Ephemeroptera), and was strongly correlated ($p < .001$) to dispersal potential (Table S5, Figure 4a). In contrast to this taxonomic vector, TAX-PCO4 showed the highest species scores for both blackflies and mayflies (Figure 4b), and was strongly correlated to site occupancy ($p = .007$, Table S5). Finally, TRA-PCO2 reflects the influence of functional feeding groups and body size on model performance (Figure 4c), as indicated by the strong correlation of both variables (i.e. $p < .001$) with this trait vector (Table S5).

Discussion

Single species models

Our results indicated that single species distributions of stream insects are highly variable in terms of predictability, as well as the significant environmental and spatial predictors

underlying such distributions. There was no evident association between model accuracy and particular taxonomic groups (Table S3). Nevertheless, a few generalisations can be highlighted with regard to the results obtained. For example, water temperature and shading, and to a lesser extent, stream width, cobbles and moss, were more relevant as environmental predictors of species distributions than stream flow or water chemistry variables (Table S3, Figure 3). This is in line with the well-known influence of temperature and resource availability on insect life cycles at high latitudes (Danks, 2007) and indicates the influence of species sorting processes along these environmental gradients. Resource availability is represented in our case by shading, which indicates the proximity of terrestrial vegetation and hence is a surrogate of availability of allochthonous resources from terrestrial origin for aquatic insect larvae. This typically corresponds with a situation of a low-order stream which, as in our case, is influenced strongly by terrestrial material from riparian vegetation which is then taken as food resource by shredders, hence promoting their dominance (Vannote, Minshall, Cummins, Sedell, & Cushing, 1980). Shading may also be inversely related to primary productivity, but in this study, we found that the relationship of species distribution with shading was always positive (Figure 3), suggesting that rather than biofilm production, it is the external input of terrestrial material from riparian birch tree abundance what is likely driving species distributions. In our case, shading was selected as a significant variable in binomial models for some predators (*Isoperla difformis* and *Plectrocnemia conspersa*) and shredders (*Leuctra* spp.), for some collector-gatherers (*Corynoneura lobata*-type, *Eukiefferiella devonica*-group, *Orthocladius rhyacobius*-group and *Tvetenia discoloripes*), and for some collector-filterers (*Philopotamus montanus* and *Prosimulium hirtipes*) (Table S3). These latter groups perhaps benefit indirectly from the increase in potential resources that the variable “shading” represents for shredders, for example, through the enhancement of

nutrient re-cycling by shredding coarse plant litter (Covich, Palmer, & Cowl, 1999; Wallace & Webster, 1996).

Spatial variables were also relevant for the distributions of some species. Specifically, large-scale spatial variables were more important than small-scale variables in explaining species distributions in our study (Table S3, Figure 3). At a larger spatial extent (ca. 500 km latitudinal gradient), previous findings indicate a stronger relevance of environmental factors, compared to spatial restrictions, on single-species distributions (Heino & de Mendoza, 2016). This is perhaps not surprising because increasing the spatial extent may have a strong positive effect on the relevance of niche processes through larger environmental gradients (Chase, 2014). However, increasing the spatial extent may also preclude species to reach environmentally suitable locations owing to dispersal limitation, and thus the relative contribution of both environmental and spatial constraints on species distributions does not always vary predictably with spatial scale (Alahuhta & Heino, 2013).

Comparative analysis across species

Comparative analysis across the species models showed a clear relationship between model performance and site occupancy. Specifically, the binomial GLMs that we built upon spatial variables could be related to site occupancy, and to a lesser extent, to taxonomic and trait vectors, whereas none of these variables was significantly related to model performance when models were based on local environmental or climate variables (Table 1). At first glance, our results also suggested both a slight influence of female dispersal potential (related to the taxonomic vector TAX-PCO3), and a potential influence of functional feeding groups and body size (related to the trait vector TRA-PCO2), on the performance of models based on spatial variables. The taxonomic vector TAX-PCO3 perhaps relates to female dispersal potential, as species scores along this vector were much higher for the blackflies than for the

rest of species, and lowest for the mayflies (Figure 4). Blackflies are possibly the best active dispersers among all the insects we considered, because females feed as flying adults and in most species they must actively search for blood meals, often several kilometers away from their natal streams (Baldwin et al., 1975). However, adult mayflies, do not feed and often have extremely short life spans (Brittain, 1990). Therefore, it seems reasonable to assume that blackflies may actively disperse better than mayflies. Site occupancy and dispersal potential were not correlated (Table S5), and both taxa were the ones with highest number of sites occupied (Fig. S4). In contrast, mayflies differed in site occupancy from non-biting midges (Chironomidae) (Fig. S4), despite species in both groups can be considered weak active dispersers, as chironomid adults are also short-lived and generally weak active fliers (Armitage, 1995). On the other hand, the trait vector TRA-PCO2 suggests an influence of feeding behaviour and body size (Figure 4, Table S5) on model performance. This is because the exploitation of food resource from terrestrial origin (i.e. shredders) would facilitate the development of more complex trophic food webs with the inclusion of predators (Figure 4). This would also contribute to the positive association of body size to TRA-PCO2 (Fig. 4), as the largest insects we found are either predators or shredders (Table S2).

Nevertheless, it is important to note that taxonomic and trait vectors had a comparatively much weaker effect on predictability by spatial variables than that of site occupancy. In fact, not only did site occupancy attain a higher statistical significance (Table 1), but it could also be partly related to the capability of the taxonomic vector TAX-PCO4 to account for the adjusted D^2 values of binomial GLMs because these two predictor variables were significantly correlated (Table S5). Moreover, when repeating the beta regression analysis by using only the significant variables selected (i.e. site occupancy, TAX-PCO3, TAX-PCO4, and TRA-PCO2), only site occupancy was statistically significant, indicating that the influence of taxonomic and trait vectors on model performance is rather weak.

Analysing through beta regression the univariate relationship of these variables with the adjusted D^2 values of binomial GLMs based on spatial effects again resulted in site occupancy as the only significant variable (see Results above). Therefore, we must conclude that any potential effect of taxonomic and trait vectors on model performance, including the effect of female dispersal potential and body size, and that of functional feeding groups, must be considered with caution: their statistical significance only appears after controlling for site occupancy and the other variables considered in the full model of beta regression. In this regard, the fact that *Baetis rhodani* is a widespread mayfly, which could not be modelled because it was present at all sites, also gives support to the idea that dispersal abilities are not so important in structuring invertebrate assemblages in high-latitude drainage basins. This is because it demonstrates that mayfly species can be widespread, despite being rather weak active dispersers. We also acknowledge that the rarest species (i.e. present in less than six sites) were not modelled because models based on such small number of presences were considered unreliable (e.g. Pierce & Ferrier, 2000). However, excluding these species does not undermine the conclusion that the distributions of most common species are better accounted for by models based on spatial variables than that of not-so-common species. In fact, we effectively modelled 47 out of the 86 taxa available at the species (most cases) or species-group (few cases) taxonomic resolution, comprising 55% of cases, which is a representative subset of species in the entire metacommunity.

Approaching the suitability of metacommunity analysis frameworks

With the information above about single-species distribution models and subsequent comparative analysis across species, it is possible to proceed with the evaluation of the suitability of the two different frameworks of metacommunity analysis (Figure 1) considered here: (1) the classical approach exemplified by the four different non-exclusive perspectives

described by Leibold et al. (2004) or (2) the three exclusive components as proposed by Logue et al. (2011).

Among the four different metacommunity perspectives of the Leibold et al. (2004) framework, neutral theory and patch dynamics do not rely on the effect of environmental variables, in contrast to species sorting and source-sink dynamics, the latter of which also incorporating a strong influence of spatial effects (Figure 1a). In our study, single-species models often relied on the effect of environmental variables, particularly temperature and shading, while being also dependent on large-scale spatial variables (Figure 3). As environmental and spatial factors are both relevant for the distribution of species, this result suggests that either species sorting along spatially structured environmental gradients, or source-sink dynamics between populations of high-quality and low-quality habitats, are both likely as important processes driving metacommunities. Then, the comparative analysis across species showed that site occupancy is responsible for the observed differences in the relevance of spatial variables on species distributions (Table 1). This suggests that common species would be better able than rare species to maintain populations in low-quality habitats through constant immigration, favouring the source-sink dynamics perspective over species sorting.

Although species-sorting processes cannot be completely discarded because of the demonstrated influence of environmental variables in many cases, deviance partitioning suggests that the pure effects of environmental and spatial factors on species distributions are stronger than their joint effects (Table S3). Also, the effect of spatial variables was better explained than that of environmental factors by our explanatory variables, particularly site occupancy, in the comparative analysis. These results slightly undermine the idea of species-sorting across spatially structured environmental gradients as the most important process shaping metacommunities. In any case, the neutral theory, which relies entirely on spatial

dynamics, is unlikely. As the dispersal potential of species has a rather weak effect on model accuracy, patch dynamics can be discarded as well as a suitable perspective of metacommunity analysis in our case. It should be acknowledged, however, that the difficulty to explain model performance with dispersal ability can also be a consequence of the coarseness of the dispersal measures currently available for freshwater invertebrates (Schmidt-Kloiber & Hering, 2015; Serra et al., 2016; Tachet et al., 2010). Moreover, the different metacommunity paradigms from Leibold et al. (2004) may always act simultaneously to a certain extent along a continuum (Figure 1a) rather than being distinct and mutually exclusive options (Brown, Sokol, et al., 2017; Gravel, Canham, Beaudet, & Messier, 2006; Logue et al., 2011).

Spatial autocorrelation may appear not only as a consequence of mass effects or species sorting along spatially structured environmental gradients when the spatial scale is not very large, but also as a consequence of dispersal limitation at very large spatial scales (Heino et al., 2015). Nevertheless, some insect species found in this study exemplify well the potential importance of the source-sink dynamics for metacommunities in subarctic streams, independently of their dispersal capability. For example, six blackfly species were examined (Table S1), of which five were present in more than 50% of sites, three of them in 75% of sites or more (Table S2). Thus, blackfly species in subarctic streams have successfully spread widely, which is advantageous to maintain metapopulations through source-sink dynamics. On the other hand, the mayflies are as widespread as the blackflies (Fig. S4), but far less capable of active dispersal. This suggests that the dispersal capability of species does not determine the metapopulation dynamics, whereas site occupancy probably does so. Spatial autocorrelation patterns have been described for the blackflies at small spatial scales, driven by strong effects of inter-specific competition for oviposition sites, and subsequent priority effects at the community level (McCreadie & Adler, 2012). The importance of priority effects

for the blackflies reinforces the idea of the relevance of site occupancy for community dynamics, where rare species are in clear disadvantage for habitat recolonisation.

Alternative to the framework of Leibold et al. (2004), we can interpret our results under the framework of Logue et al. (2011), whereby three different and mutually exclusive components can be used to analyse metacommunities: species equivalence, habitat heterogeneity and dispersal (Figure 1b). In our case, this alternative framework makes interpretation of the results much easier. At the very least, we can conclude that species equivalence is unlikely to play any role in metacommunity dynamics, similarly to discarding neutral theory under the Leibold *et al.* (2004) framework. Dispersal can also be discarded, yet again with caution due to the current lack of high resolution dispersal measures for freshwater invertebrates (Schmidt-Kloiber & Hering, 2015; Serra et al., 2016; Tachet et al., 2010). Thus, the main difference in the interpretation of the results with this alternative framework is that we can now be certain about the role of habitat heterogeneity, while under the Leibold et al. (2004) framework it is more difficult to discern whether species sorting or source-sink dynamics is the dominant process. Habitat heterogeneity is indeed related to both mechanisms. In fact, using habitat heterogeneity in space and time as the templet for ecological strategies (Southwood, 1977) could be the framework of choice in situations where it is difficult to discern species sorting processes from source-sink dynamics.

Alternative approaches, caveats and conclusions

Emergent properties at the community level are difficult to discern from field observational data alone. In this regard, population genetics can be very useful in order to gain confidence about the distinction between, e.g., source-sink dynamics and species sorting processes. This is because population genetic studies could be used to estimate the relative contribution of immigrants from nearby populations to the genetic variability of the population under study

(Bunn & Hughes, 1997; Hughes, Huey, & Schmidt, 2013; Hughes, Schmidt, & Finn, 2009). Genetic analyses would probably provide the opportunity for a more robust interpretation of our results. Genetic studies, however, are difficult to accomplish with stream insects in the field when the idea is to compare many species at a time, and they are far more expensive than the comparative approach of single species distributions we considered here. Therefore, the comparative approach presented here can be used as a first step to explore the relative contribution of environmental and spatial factors on species distributions, without using expensive and time-consuming genetic analyses. In fact, by using the comparative approach we can certainly conclude that the dispersal capability of species and neutral theory play little role in shaping subarctic stream insect metacommunities. Rather, it is habitat heterogeneity, which influences mass effects and/or species sorting processes, that matters. Subsequently, the results of our study strongly recommend the preservation of habitat heterogeneity as the conservation strategy to maintain biodiversity in these ecosystems.

Nevertheless, it should be acknowledged that one shortcoming of single-species distribution modelling is that it does not consider the influence of species interactions in structuring ecological communities. Stream ecology has considered that severe environmental conditions may weaken the potential effects of biotic interactions in structuring communities (Peckarsky, 1983). However, more recent findings pose doubts as to whether this is actually true (Cadotte & Tucker, 2017; Thomson et al., 2002). In fact, biotic interactions can reproduce patterns of community structure essentially identical to what it could be expected from environmental filtering alone. This is because environmental changes may affect population growth rates of competing species in opposite ways, and this may cause the exclusion of some species that would otherwise be able to coexist (Cadotte & Tucker, 2017). There exists also evidence indicating that biotic interactions limit the geographical range expansion of species facing environmental changes (Pigot & Tobias, 2013; Sexton et al.,

2009). Overall, this suggests that inter-specific interactions may also play a role in our case, although the abundances of insect larvae in subarctic streams are typically low (see also Heino & Grönroos, 2017) and may thus result in weak density-dependent interactions among species (see also Morin, 2011).

Our study considered tributary streams draining into two linear sub-elements of a larger river network (Fig. S1). However, there exists growing concern about the potential role of the entire dendritic river networks in shaping biodiversity patterns, community structure and species distributions (Altermatt, 2013; Altermatt & Fronhofer, 2017; Brown, Wahl, & Swan, 2017; Jamoneau, Passy, Soininen, Lebuocher, & Tison-Rosebery, 2017; Schmera et al., 2017). For example, the consideration of whole river networks may unveil a more preeminent role for spatial factors in community assembly, undermining the role of environmental filtering. Therefore, studies conducted across whole dendritic networks could be more in line with neutral theory, as shown by Muneeppeerakul et al. (2008) for fish communities, yet no environmental variable was truly considered in that study. Although we focused on tributary streams draining into the main river, the consideration of whole dendritic networks may help us to perceive more accurately the real connectivity pathways between isolated patches. This connectivity may have consequences for metacommunity stability with respect to a situation where only a linear component of this network is acting (Fagan, 2002). Also, dispersal along dendritic networks implies more variability in local richness with strong consequences also for community differentiation among patches (Carrara, Altermatt, Rodriguez-Iturbe, & Rinaldo, 2012; Seymour, Fronhofer, & Altermatt, 2015). In any case, dispersal processes in stream networks may depend on the organism group considered (Schmera et al. 2017).

In the case of stream insects, the taxa considered and the taxonomic resolution achieved prior to species-distribution modelling, may also have important consequences on

our perception of the influence of dendritic riverine networks on biodiversity patterns (Kaelin & Altermatt, 2016). Here, some taxa were discarded as it was not possible to determine the species. Provided that the influences of dendritic landscapes and biotic interactions (discussed above) may strongly affect how we understand the reality of community assemblages, it is essential to use the best taxonomic resolution possible to make accurate inferences about the mechanisms truly governing the observed patterns. In fact, the criterion of ‘best taxonomic resolution possible’ used in our modelling endeavours is a fundamental requirement to draw robust conclusions to be applied in biodiversity conservation.

Finally, for biodiversity conservation, it is essential to focus on maintaining habitat heterogeneity because it appears to determine metacommunity organization (Kärnä et al. 2015) and species distributions (Heino & de Mendoza, 2016) in streams at high latitudes. Unless habitat heterogeneity is not considered (along with potentially important effects of dendritic network structure and biotic interactions), conservation plans may fall short and not result in desired outcomes.

Acknowledgements

We thank Sirkku Lehtinen and Marja Lindholm for help in the field or the laboratory. We also thank two anonymous reviewers and the Guest Editor Florian Altermatt for their very helpful and constructive comments on an earlier version of this manuscript. Kevo Subarctic Research Station provided facilities during the field work. This study is part of the project “Spatial scaling, metacommunity structure and patterns in stream communities” that was supported financially by a grant from the Academy of Finland. Further support was provided by grants (no: 273557, no: 267995 and no: 285040) from the Academy of Finland. The authors declare no conflict of interest.

References

- Aalto, J., Pirinen, P., Heikkinen, J., & Venäläinen, A. (2013) Spatial interpolation of monthly climate data for Finland: comparing the performance of kriging and generalized additive models. *Theoretical and Applied Climatology*, **112**, 99–111.
- Adler, P.H., & Crosskey, R.W. (2016) *World blackflies (Diptera: Simuliidae): A comprehensive revision of the taxonomic and geographical inventory*. Retrieved from <http://www.clemson.edu/cafls/biomia/pdfs/blackflyinventory.pdf>.
- Alahuhta, J., & Heino, J. (2013) Spatial extent, regional specificity and metacommunity structuring in lake macrophytes. *Journal of Biogeography*, **40**, 1572–1582.
- Altermatt, F. (2013) Diversity in riverine metacommunities: a network perspective. *Aquatic Ecology*, **47**, 365–377.
- Altermatt, F., & Fronhofer, E. A. (2017) Dispersal in dendritic networks: Ecological consequences on the spatial distribution of population densities. *Freshwater Biology*, **63**, 22–32. <https://doi.org/10.1111/fwb.12951>
- Armitage, P. D. (1995) Behaviour and ecology of adults. In P. D. Armitage, P. S. Cranston, & L. C. V. Pinder (Eds.), *The Chironomidae: Biology and Ecology of Non-Biting Midges* (pp. 194–224). London: Chapman & Hall.
- Baldwin, W. F., West, A. S., & Gomery, J. (1975) Dispersal pattern of black flies (Diptera: Simuliidae) tagged with ³²P. *The Canadian Entomologist*, **107**, 113–118.
- Bonthoux, S., Baselga, A., & Balent, G. (2013) Assessing community-level and single-species models predictions of species distributions and assemblage composition after 25 years of land cover change. *PLoS ONE*, **8**, e54179.
- Borcard, D., Gillet, F., & Legendre, P. (2011) *Numerical Ecology with R*. New York: Springer.

659 Brittain, J. E. (1990) Life history strategies in Ephemeroptera and Plecoptera. In I. C.
660 Campbell (Ed.), *Mayflies and Stoneflies* (pp. 1–12). Dordrecht: Kluwer Academic
661 Publishers.

662 Brown, B. L., Sokol, E. R., Skelton, J., & Tornwall, B. (2017) Making sense of
663 metacommunities: dispelling the mythology of a metacommunity typology.
664 *Oecologia*, **183**, 643–652.

665 Brown, B. L., Wahl, C., & Swan, C. M. (2017). Experimentally disentangling the influence
666 of dispersal and habitat filtering on benthic invertebrate community structure.
667 *Freshwater Biology*, **63**, 48–61. <https://doi.org/10.1111/fwb.12995>

668 Bunn, S. E., & Hughes, J. M. (1997) Dispersal and recruitment in streams: evidence from
669 genetic studies. *Journal of the North American Benthological Society*, **16**, 338–346.

670 Burnham, K. P., & Anderson, D. R. (2004) Multimodel inference: understanding AIC and
671 BIC in model selection. *Sociological Methods & Research*, **33**, 261–304.

672 Cadotte, M. W., & Tucker, C. M. (2017) Should environmental filtering be abandoned?
673 *Trends in Ecology & Evolution*, **32**, 429–437.

674 Carrara, F., Altermatt, F., Rodriguez-Iturbe, I., & Rinaldo, A. (2012) Dendritic connectivity
675 controls biodiversity patterns in experimental metacommunities. *Proceedings of the*
676 *National Academy of Sciences of the United States of America*, **109**, 5761–5766.

677 Chapman, D. S., & Purse, B. V. (2011) Community versus single-species distribution models
678 for British plants. *Journal of Biogeography*, **38**, 1524–1535.

679 Chase, J. M. (2014) Spatial scale resolves the niche versus neutral theory debate. *Journal of*
680 *Vegetation Science*, **25**, 319–322.

681 Cook, R. D. (1977) Detection of influential observation in linear regression. *Technometrics*,
682 **19**, 15–18.

683 Covich, A. P., Palmer, M. A., & Cowl, T. A. (1999) The role of benthic invertebrate species
684 in freshwater ecosystems: Zoobenthic species influence energy flows and nutrient
685 cycling. *BioScience*, **49**, 119-127.

686 Cribari-Neto, F., & Zeileis, A. (2010) Beta regression in R. *Journal of Statistical Software*,
687 **34**, 1–24.

688 Crosskey, R. W. (1990) *The Natural History of Blackflies*. John Wiley & Sons, Chichester.

689 Dankers, R., & Christensen, O.B. (2005) Climate change impacts on snow coverage,
690 evaporation and river discharge in the sub-arctic Tana basin, Northern Fennoscandia.
691 *Climatic Change*, **69**, 367–392.

692 Danks, H. V. (2007) How aquatic insects live in cold climates. *The Canadian Entomologist*,
693 **139**, 443–471.

694 de Jong, Y., Verbeek, M., Michelsen, V., Bjørn, P. P., Los, W., Steeman, F., Bailly, N.,
695 Basire, C., Chylarecki, P., Stloukal, E., Hagedorn, G., Wetzel, F. T., Glöckler, F.,
696 Kroupa, A., Korb, G., Hoffmann, A., Häuser, C., Kohlbecker, A., Müller, A.,
697 Güntsch, A., Stoev, P., & Penev, L. (2014) Fauna Europaea: All animal species on the
698 web. *Biodiversity Data Journal*, **2**, e4034.

699 Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Marquéz, J.
700 R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne,
701 P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S.
702 (2013) Collinearity: A review of methods to deal with it and a simulation study
703 evaluating their performance. *Ecography*, **36**, 27–46.

704 Dray, S., Legendre, P., & Peres-Neto P. R. (2006) Spatial modelling: A comprehensive
705 framework for principal coordinate analysis of neighbour matrices (PCNM).
706 *Ecological Modelling*, **196**, 483–493.

707 Fagan, W. F. (2002) Connectivity, fragmentation, and extinction risk in dendritic
708 metapopulations. *Ecology*, **83**, 3243–3249.

709 Ferrari, S. L. P., & Cribari-Neto, F. (2004). Beta regression for modelling rates and
710 proportions. *Journal of Applied Statistics*, **31**, 799–815.

711 Ferrier, S., & Guisan, A. (2006) Spatial modelling of biodiversity at the community level.
712 *Journal of Applied Ecology*, **43**, 393–404.

713 Fox, J. (2005) The R Commander: A Basic Statistics Graphical User Interface to R. *Journal*
714 *of Statistical Software*, **14(9)**, 1–42.

715 Gíslason, G. M., Hannesdóttir, E. R., Munoz, S. S., & Pálsson, S. (2015) Origin and dispersal
716 of *Potamophylax cingulatus* (Trichoptera: Limnephilidae) in Iceland. *Freshwater*
717 *Biology*, **60**, 387–394.

718 Goslee, S.C. & Urban, D.L. (2007) The ecodist package for dissimilarity-based analysis of
719 ecological data. *Journal of Statistical Software*, **22(7)**, 1–19.

720 Gravel, D., Canham, C. D., Beaudet, M., & Messier, C. (2006) Reconciling niche and
721 neutrality: The continuum hypothesis. *Ecology Letters*, **9**, 399–409.

722 Grönroos, M., Heino, J., Siqueira, T., Landeiro, V. L., Kotanen, J., & Bini, L. M. (2013)
723 Metacommunity structuring in stream networks: Roles of dispersal mode, distance
724 type, and regional environmental context. *Ecology and Evolution*, **3**, 4473–4487.

725 Guisan, A., & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology.
726 *Ecological Modelling*, **135**, 147–186.

727 Hanski, I. (1994) A practical model of metapopulation dynamics. *Journal of Animal Ecology*,
728 **63**, 151–162.

729 Heino, J. (2005) Positive relationship between regional distribution and local abundance in
730 stream insects: a consequence of niche breadth or niche position? *Ecography*, **28**,
731 345–354.

732 Heino, J., & de Mendoza, G. (2016) Predictability of stream insect distributions is dependent
 733 on niche position, but not on biological traits or taxonomic relatedness of species.
 734 *Ecography*, **39**, 1216–1226.

735 Heino, J., & Grönroos, M. (2014) Untangling the relationships among regional occupancy,
 736 species traits and niche characteristics in stream invertebrates. *Ecology and Evolution*,
 737 **4**, 1931–1942.

738 Heino, J., & Grönroos, M. (2017) Exploring species and site contributions to beta diversity in
 739 stream insect assemblages. *Oecologia*, **183**, 151–160.

740 Heino, J., Ilmonen, J., & Paasivirta, L. (2014) Continuous variation of macroinvertebrate
 741 communities along environmental gradients in northern streams. *Boreal*
 742 *Environment Research*, **19**, 21–38.

743 Heino, J., Melo, A. S., Siqueira, T., Soininen, J., Valanko, S., & Bini, L. M. (2015)
 744 Metacommunity organisation, spatial extent and dispersal in aquatic systems:
 745 Patterns, processes and prospects. *Freshwater Biology*, **60**, 845–869.

746 Hoffsten, P.-O. (2004) Site-occupancy in relation to flight-morphology in caddisflies.
 747 *Freshwater Biology*, **49**, 810–817.

748 Hubbell, S. P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*.
 749 Princeton, NJ: Princeton University Press.

750 Hughes, J. M., Huey, J. A. & Schmidt, D. J. (2013) Is realised connectivity among
 751 populations of aquatic fauna predictable from potential connectivity? *Freshwater*
 752 *Biology*, **58**, 951–966.

753 Hughes, J. M., Schmidt, D. J., & Finn, D.S. (2009) Genes in streams: Using DNA to
 754 understand the movement of freshwater fauna and their riverine habitat. *BioScience*,
 755 **59**, 573–583.

756 Ilmonen, J. (2014) Checklist of the family Simuliidae (Diptera) of Finland. *ZooKeys*, **441**,
757 91–95.

758 Jamoneau, A., Passy, S. I., Soininen, J., Lebuocher, T., & Tison-Rosebery, J. (2017) Beta
759 diversity of diatom species and ecological guilds: Response to environmental and
760 spatial mechanisms along the stream watercourse. *Freshwater Biology*, **63**, 62–73.
761 <https://doi.org/10.1111/fwb.12980>

762 Kaelin, K., & Altermatt, F. (2016) Landscape-level predictions of diversity in river networks
763 reveal opposing patterns for different groups of macroinvertebrates. *Aquatic
764 Ecology*, **50**, 283–295.

765 Kärnä, O.-M., Grönroos, M., Antikainen, H., Hjort, J., Ilmonen, J., Paasivirta, L., & Heino, J.
766 (2015) Inferring the effects of potential dispersal routes on the metacommunity
767 structure of stream insects: as the crow flies, as the fish swims or as the fox runs?
768 *Journal of Animal Ecology*, **84**, 1342–1353.

769 Legendre, P., & Legendre L. (2012) *Numerical Ecology (3rd English edition)*. Elsevier,
770 Amsterdam.

771 Legendre, P., Borcard, D., Blanchet, F.G., & Dray, S. (2013). *PCNM: MEM spatial
772 eigenfunction and principal coordinate analyses*. R package version 2.1-2/r109.
773 Retrieved from <http://R-Forge.R-project.org/projects/sedar/>

774 Leibold, M. A. (1995) The niche concept revisited: mechanistic models and community
775 context. *Ecology*, **76**, 1371–1382.

776 Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F.,
777 Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M., & Gonzalez, A. (2004)
778 The metacommunity concept: a framework for multi-scale community ecology.
779 *Ecology Letters*, **7**, 601–613.

780 Logue, J. B., Mouquet, N., Peter, H., & Hillebrand, H. (2011) Empirical approaches to
 781 metacommunities: a review and comparison with theory. *Trends in Ecology and*
 782 *Evolution*, **26**, 482–491.

783 Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2013) *cluster: Cluster*
 784 *analysis basics and extensions*. R package version 1.14.4. Retrieved from
 785 <http://CRAN.R-project.org/package=cluster>.

786 Mansikkaniemi, H. (1970) Deposits of sorted material in the Inarijoki-Tana river valley in
 787 Lapland. *Reports of Kevo Subarctic Research Station*, **6**, 1–63.

788 McCreddie, J. W., & Adler, P. H. (2012) The roles of abiotic factors, dispersal, and species
 789 interactions in structuring stream assemblages of black flies (Diptera: Simuliidae).
 790 *Aquatic Biosystems*, **8**, 14.

791 Merritt, R. W., & Cummins, K. W. (1996) *An Introduction to the Aquatic Insects of North*
 792 *America*, 3rd ed. Dubuque: Kendall/Hunt Publishing.

793 Morin, P. J. (2011) *Community Ecology*, 2nd ed. Oxford: Wiley-Blackwell.

794 Müller-Peddinghaus, E. (2011) *Flight-morphology of Central European caddisflies (Insecta:*
 795 *Trichoptera) in relation to their ecological preferences*. PhD Thesis, University of
 796 Duisburg-Essen, Essen.

797 Müller-Peddinghaus, E., & Hering, D. (2013) The wing morphology of limnephilid
 798 caddisflies in relation to their habitat preferences. *Freshwater Biology*, **58**, 1138–
 799 1148.

800 Muneeppeerakul, R., Bertuzzo, E., Lynch, H. J., Fagan, W. F., Rinaldo, A., & Rodriguez-
 801 Iturbe, I. (2008) Neutral metacommunity models predict fish diversity patterns in
 802 Mississippi-Missouri basin. *Nature*, **453**, 220–222.

803 National Board of Waters (1981) *Vesihallinnon analyysimenetelmät*, Tiedotus 213. Helsinki:
804 Vesihallitus.

805 Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B.,
806 Simpson, G. L., Solymos, P., Stevens, M. H. H., & Wagner, H. (2013) *vegan*:
807 *Community Ecology Package*. R package version 2.0-9. Retrieved from
808 <http://CRAN.R-project.org/package=vegan>.

809 Pearce, J., & Ferrier, S. (2000) An evaluation of alternative algorithms for fitting species
810 distribution models using logistic regression. *Ecological Modelling*, **128**, 127–147.

811 Peckarsky, B. L. (1983) Biotic interactions or abiotic limitations? A model of lotic
812 community structure. In T. D. Fontaine III, & S. M. Bartell (Eds.), *Dynamics of Lotic*
813 *Ecosystems* (pp. 303–323). Ann Arbor: Ann Arbor Science Publishers.

814 Pigot, A. L., & Tobias, J. A. (2013) Species interactions constrain geographic range
815 expansion over evolutionary time. *Ecology Letters*, **16**, 330–338.

816 Pirinen, P., Simola, H., Aalto, J., Kaukoranta, J. P., Karlsson, P., & Ruuhela, R. (2012)
817 Climatological statistics of Finland 1981–2010. *Finnish Meteorological Institute*
818 *Reports*, **1**, 1–96.

819 Pulido, C., Riera, J. L., Ballesteros, E., Chappuis, E., & Gacia, E. (2015) Predicting aquatic
820 macrophyte occurrence in soft-water oligotrophic lakes (Pyrenees mountain range).
821 *Journal of Limnology*, **74**, 143–154.

822 Pulliam, H. R. (1988) Sources, sinks, and population regulation. *American Naturalist*, **132**,
823 652–661.

824 R Core Team (2013) *R: A language and environment for statistical computing*. R Foundation
825 *for Statistical Computing*, Vienna, Austria. Retrieved from <http://www.R-project.org/>.

826 Schmera, D., Árva, D., Boda, P., Bódis, E., Bolgovics, Á., Borics, G., Cserecsa, A., Deák, C.,
827 Krasznai, E. Á., Lukács, B. A., Mauchart, P., Móra, P., Sály, P., Specziár, A., Süveges

828 K., Szivák, I., Takács, P., Tóth, M., Várbíró, G., Vojtkó, A. E., & Erős, T. (2017).
829 Does isolation influence the relative role of environmental and dispersal-related
830 processes in stream networks? An empirical test of the network position hypothesis
831 using multiple taxa. *Freshwater Biology*, **63**, 74–85.
832 <https://doi.org/10.1111/fwb.12973>

833 Schmidt-Kloiber, A., & Hering, D. (2015) www.freshwaterecology.info – An online tool that
834 unifies, standardises and codifies more than 20,000 European freshwater organisms
835 and their ecological preferences. *Ecological Indicators*, **53**, 271–282.

836 Serra, S. R. Q., Cobo, F., Graça, M. A. S., Dolédec, S., & Feio M. J. (2016) Synthesising the
837 trait information of European Chironomidae (Insecta: Diptera): Towards a new
838 database. *Ecological Indicators*, **61**, 282–292.

839 Sexton, J. P., McIntyre, P. J., Angert, A. L., & Rice, K. J. (2009) Evolution and ecology of
840 species range limits. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 415–
841 436.

842 Seymour, M., Fronhofer, E. A., & Altermatt, F. (2015) Dendritic network structure and
843 dispersal affect temporal dynamics of diversity and species persistence. *Oikos*, **124**,
844 908–916.

845 Southwood, T. R. E. (1977) Habitat, the templet for ecological strategies? *Journal of Animal*
846 *Ecology*, **46**, 337–365.

847 Tachet, H., Richoux, P., Bournaud, M., & Usseglio-Polatera, P. (2010) *Invertébrés d’eau*
848 *douce: systématique, biologie, écologie (nouvelle édition revue et augmentée)*.
849 Paris: CNRS.

850 Thomson, J. R., Lake, P. S., & Downes, B. J. (2002) The effect of hydrological disturbance
851 on the impact of a benthic invertebrate predator. *Ecology*, **83**, 628–642.

852 Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R., & Cushing, C.E. (1980) The
853 river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences*, **37**,
854 130–137.

855 Wallace, J. B., & Webster, J. R. (1996) The role of macroinvertebrates in stream ecosystem
856 function. *Annual Review of Entomology*, **41**, 115–139.

857 Wentworth, C. K. (1922) A scale of grade and class terms for clastic sediments. *Journal of*
858 *Geology*, **30**, 377–392.

859 Zar, J. H. (1984) *Biostatistical Analysis, Second Edition*. Prentice Hall, Englewood Cliffs.
860

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Table S1. Insect species considered.

Table S2. Species traits considered and site occupancy.

Table S3. Results of binomial GLMs.

Table S4. Taxonomic and trait vectors from Principal Coordinate Analysis (PCO), with corresponding scores for each species.

Table S5. Statistical significance of the correlations among site occupancy, species traits, trait vectors, and taxonomic vectors.

Figure S1. A map of the study area located in the Tenojoki drainage basin.

Figure S2. Eigenvalues from taxonomic Principal Coordinate Analysis (PCO) based on taxonomic distances between species.

Figure S3. Results of binomial GLMs in relation to site occupancy and species traits.

Figure S4. Comparison of site occupancy values between different insect groups.

876 **Table 1.** Results of beta regression showing the effects of site occupancy, biological trait
877 vectors and taxonomic vectors on different fractions of variation (adjusted deviance, Adj. D^2)
878 explained by binomial GLMs: local environment (E) effects, climate (C) effects, spatial (S)
879 effects, and combined (E+C+S) effects. Significant values ($p < .05$) are shown in boldface;
880 marginally significant values ($p < .10$) in italics.

Adj. D^2 of E effects						
	Estimate	SE	z	p	Log-likelihood	Pseudo R^2
(Intercept)	-1.6002	0.2220	-7.209	<0.001	49.06	0.2500
Site occupancy	-0.0042	0.0091	-0.469	0.646		
TAX-PCO1	-378.2072	323.7056	-1.168	0.243		
TAX-PCO2	91.4160	161.7362	0.565	0.572		
TAX-PCO3	225.8850	218.0393	1.036	0.300		
TAX-PCO4	-112.1618	101.9824	-1.100	0.271		
TRA-PCO1	-0.5557	0.8628	-0.644	0.520		
TRA-PCO2	1.3525	1.0712	1.263	0.207		
TRA-PCO3	0.3959	0.9475	0.418	0.676		
TRA-PCO4	-1.3110	1.3459	-0.974	0.330		
Adj. D^2 of C effects						
	Estimate	SE	z	p	Log-likelihood	Pseudo R^2
(Intercept)	-2.9806	0.2682	-11.115	<0.001	94.66	0.2309
Site occupancy	0.0029	0.0103	0.278	0.781		
TAX-PCO1	-579.9412	375.5543	-1.544	0.123		
TAX-PCO2	83.1028	185.9703	0.447	0.655		
TAX-PCO3	352.6843	252.8857	1.395	0.163		
TAX-PCO4	-1.9501	120.4650	-0.016	0.987		
TRA-PCO1	-1.6362	1.0045	-1.629	0.103		
TRA-PCO2	0.4327	1.2476	0.347	0.729		
TRA-PCO3	0.6454	1.1129	0.580	0.562		
TRA-PCO4	-2.1768	1.5519	-1.403	0.161		
Adj. D^2 of S effects						
	Estimate	SE	z	p	Log-likelihood	Pseudo R^2
(Intercept)	-2.4436	0.2222	-10.997	<0.001	60.19	0.2137
Site occupancy	0.0208	0.0083	2.498	0.012		
TAX-PCO1	309.6453	360.2543	0.860	0.390		
TAX-PCO2	-4.0452	162.3764	-0.025	0.980		
TAX-PCO3	365.5554	205.9633	1.775	<i>0.076</i>		
TAX-PCO4	217.5296	102.0912	2.131	0.033		
TRA-PCO1	-1.0051	0.8333	-1.206	0.228		

TRA-PCO2	2.5228	1.1999	2.102	0.036
TRA-PCO3	1.4151	0.9765	1.449	0.147
TRA-PCO4	-0.7984	1.3342	-0.598	0.550

Adj. D² of E+C+S effects

	Estimate	SE	z	<i>p</i>	Log-likelihood	Pseudo <i>R</i> ²
(Intercept)	-0.7133	0.2331	-3.060	0.002	17.76	0.1791
Site occupancy	0.0111	0.0095	1.164	0.244		
TAX-PCO1	-140.2227	352.1883	-0.398	0.691		
TAX-PCO2	32.5012	178.6572	0.182	0.856		
TAX-PCO3	407.2860	228.4095	1.783	0.075		
TAX-PCO4	81.0016	110.2921	0.734	0.463		
TRA-PCO1	-1.3052	0.9142	-1.428	0.153		
TRA-PCO2	1.6446	1.1935	1.378	0.168		
TRA-PCO3	0.9742	1.0367	0.940	0.347		
TRA-PCO4	-1.0409	1.4575	-0.714	0.475		

Figure legends

Figure 1 Conceptual representation of (a) the four non-exclusive classical approaches in metacommunity studies (Leibold et al. 2004), and (b) the more recent framework of metacommunity analysis based on three exclusive components (Logue et al. 2011); according to the relative relevance of the variables used in this study (axes): spatial variables (x-axis), environmental variables (y-axis) and the different dispersal capability of species (z-axis). Circles represent the theoretical location where the emphasis of each approach is situated across the three axes.

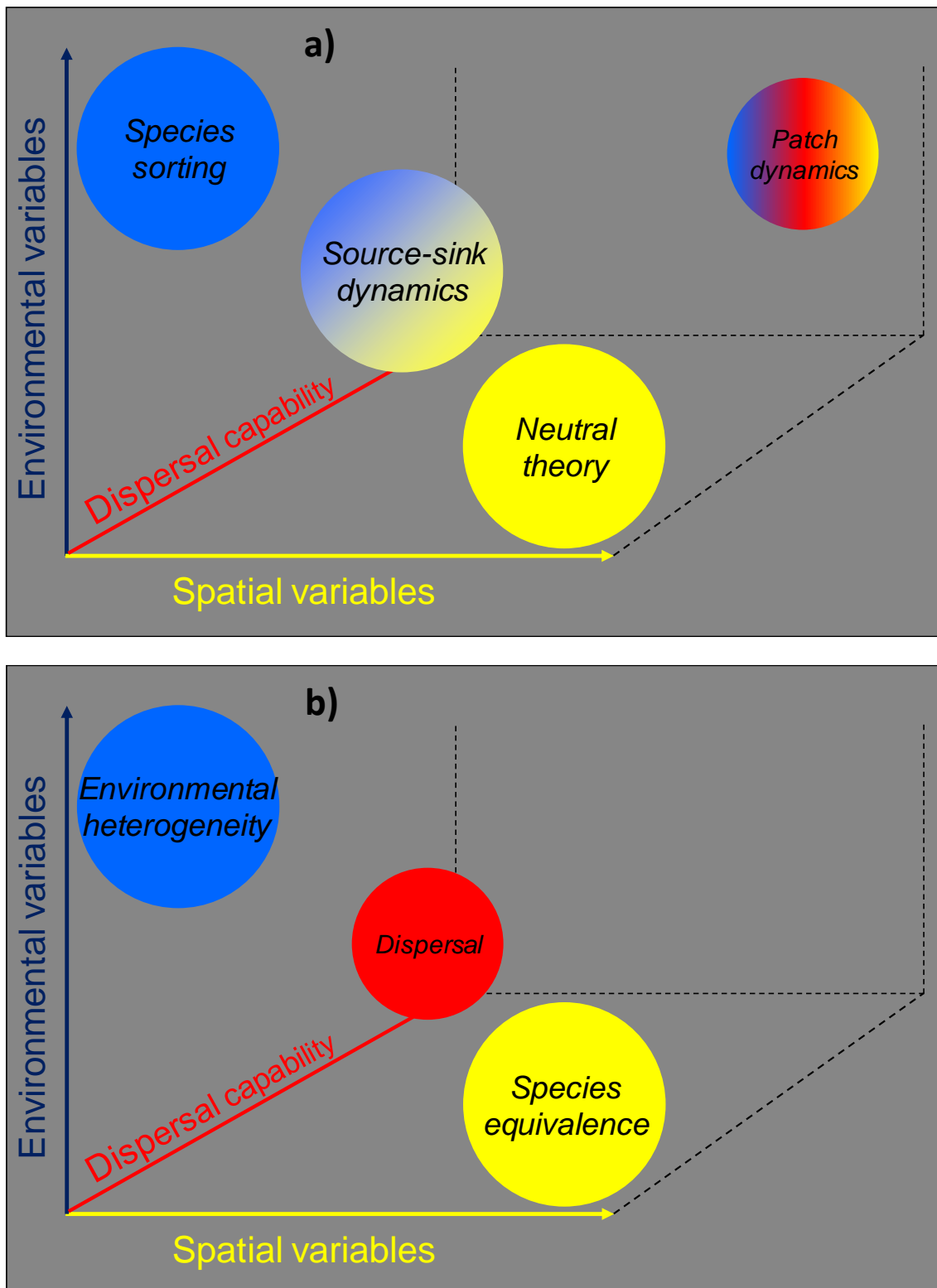
Figure 2 Flow chart of the statistical analyses performed in this study.

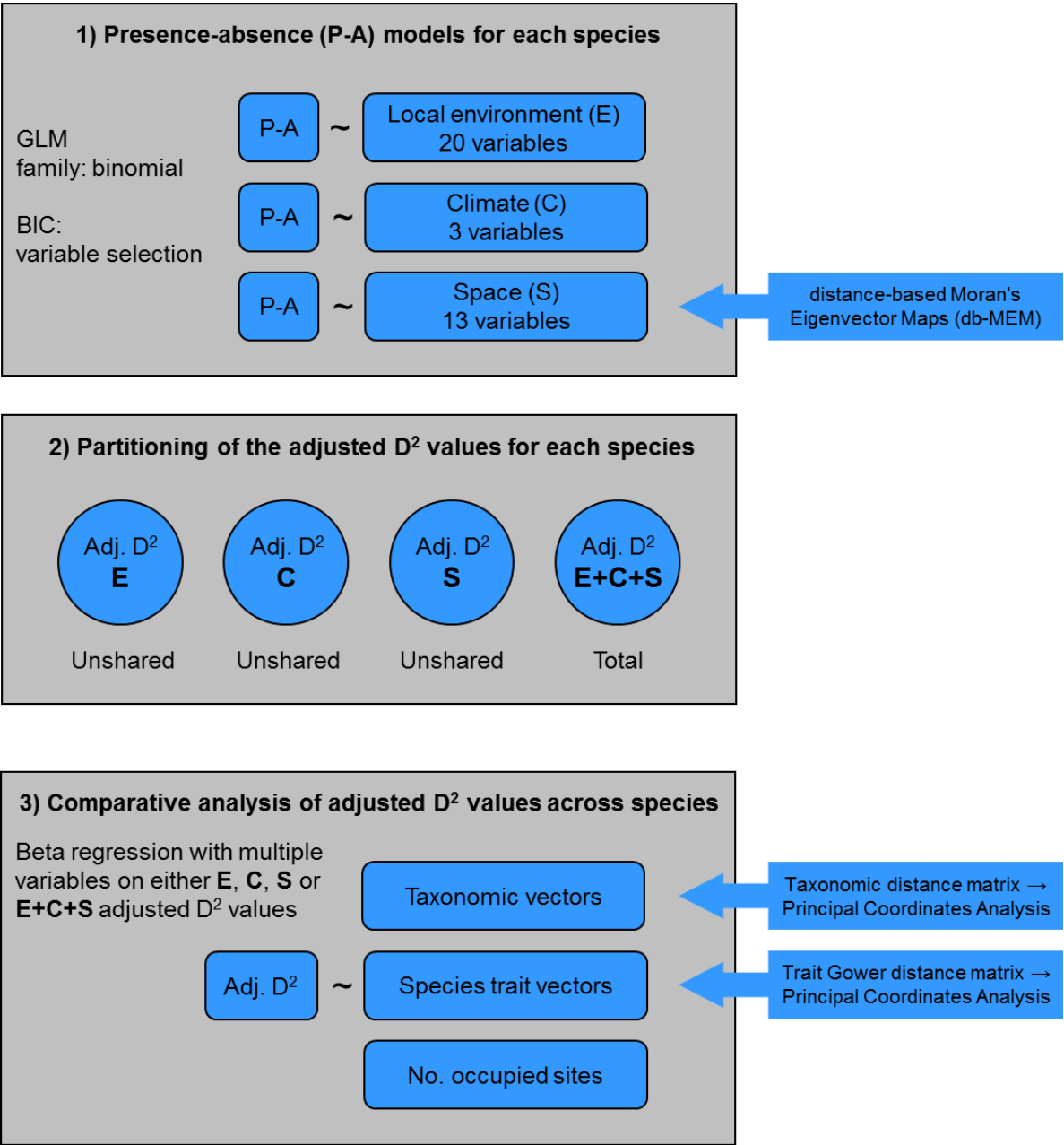
Figure 3 Frequency of local environmental variables (top) and spatial variables (bottom) selected as significant in explaining species distributions through binomial GLMs. Spatial variables are arranged from small-scale (i.e. V13) to large-scale extent (i.e. V1), and include V11 which was never selected. Climate variables are not shown (mean annual temperature was selected 32 times, and July air temperature 20 times). Species-environment relationships are shown in black when positive and grey when negative. Specific information for each species can be found in Table S2.

Figure 4 Species scores on taxonomic vectors TAX-PCO3 (a) and TAX-PCO4 (b), and on trait vector TRA-PCO2 (c), arranged from lowest to highest values. For the trait vector, the location of the different functional feeding (FFG) and habit trait groups (HTG), is indicated, as well as the four different body size classes considered (BS, represented by columns of four different sizes), and those insects considered as of high female dispersal potential (H).

904 **Figure 1**

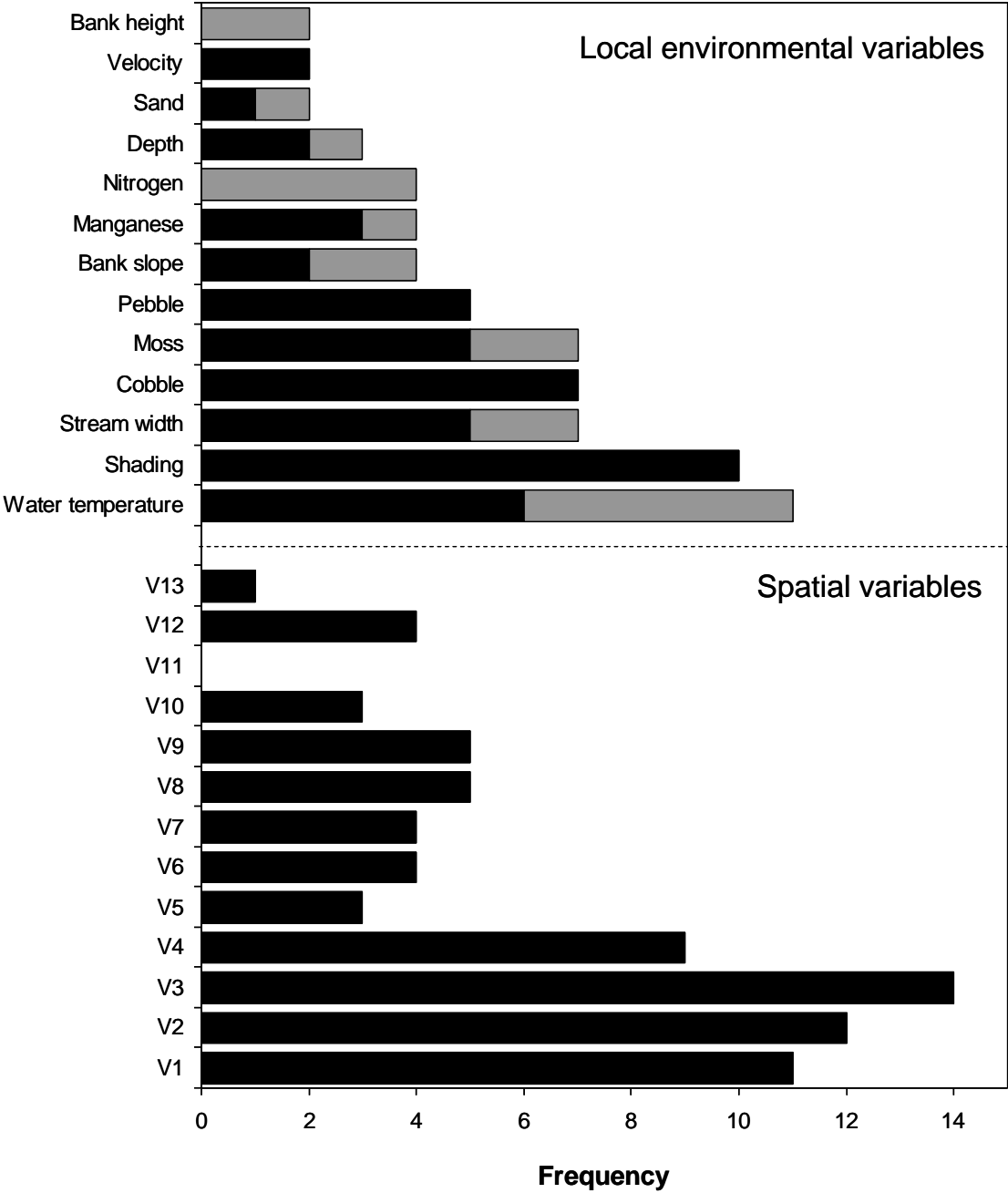
905

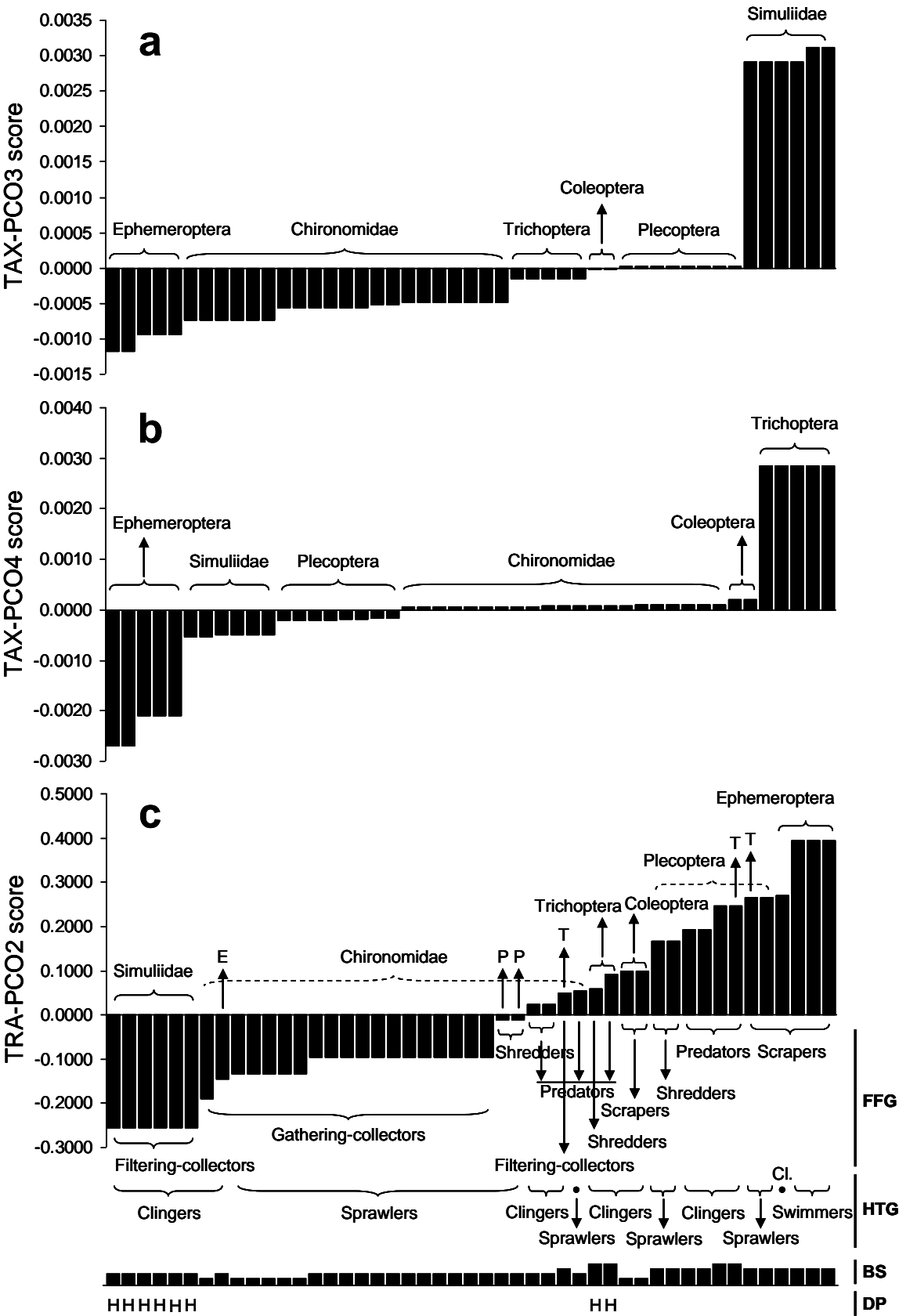




908 **Figure 3**

909





Supporting Information

Highly variable species distribution models in a subarctic stream metacommunity: patterns, mechanisms and implications

Guillermo de Mendoza, Riikka Kaivosoja, Mira Grönroos, Jan Hjort, Jari Ilmonen,
Olli-Matti Kärnä, Lauri Paasivirta, Laura Tokola & Jani Heino

Freshwater Biology

Table S1. Insect species considered.

Table S2. Species traits considered and site occupancy.

Table S3. Results of binomial GLMs.

Table S4. Taxonomic and trait vectors from Principal Coordinate Analysis (PCO), with corresponding scores for each species.

Table S5. Statistical significance of the correlations among site occupancy, species traits, trait vectors, and taxonomic vectors.

Figure S1. A map of the study area located in the Tenojoki drainage basin.

Figure S2. Eigenvalues from taxonomic Principal Coordinate Analysis (PCO) based on taxonomic distances between species.

Figure S3. Results of binomial GLMs in relation to site occupancy and species traits.

Figure S4. Comparison of site occupancy values between different insect groups.

Table S1. Insect species considered, ordered alphabetically, with information about their higher classification (order, family) and codes used.

Species	Higher classification	Code
<i>Ameletus inopinatus</i> Eaton	Ephemeroptera, Ameletidae	Ame.ino
<i>Amphinemura borealis</i> (Morton)	Plecoptera, Nemouridae	Amp.bor
<i>Amphinemura sulcicollis</i> (Stephens)	Plecoptera, Nemouridae	Amp.sul
<i>Apatania muliebris</i> McLachlan	Trichoptera, Apataniidae	Apa.mul
<i>Baetis muticus</i> (Linnaeus)	Ephemeroptera, Baetidae	Bae.mut
<i>Baetis</i> cf. <i>niger</i> (Linnaeus)	Ephemeroptera, Baetidae	Bae.nig
<i>Brachyptera risi</i> (Morton)	Plecoptera, Taeniopterygidae	Bra.ris
<i>Cardiocladius capucinus</i> (Zetterstedt)	Diptera, Chironomidae	Car.cap
<i>Cardiocladius fuscus</i> Kieffer	Diptera, Chironomidae	Car.fus
<i>Constempellina brevicosta</i> (Edwards)	Diptera, Chironomidae	Con.bre
<i>Corynoneura lobata</i> -type	Diptera, Chironomidae	Cor.lob
<i>Diura nanseni</i> (Kempny)	Plecoptera, Perlodidae	Diu.nan
<i>Elmis aenea</i> (Muller)	Coleoptera, Elmidae	Elm.aen
<i>Ephemerella aroni</i> Eaton	Ephemeroptera, Ephemerellidae	Eph.aro
<i>Eukiefferiella boevrensis</i> Brundin	Diptera, Chironomidae	Euk.boe
<i>Eukiefferiella brevicar</i> (Kieffer)	Diptera, Chironomidae	Euk.bre
<i>Eukiefferiella devonica</i> -group	Diptera, Chironomidae	Euk.dev
<i>Helodon ferrugineus</i> (Wahlberg)	Diptera, Simuliidae	Hel.fer
<i>Heptagenia dalecarlica</i> Bengtsson	Ephemeroptera, Heptageniidae	Hep.dal
<i>Hydraena gracilis</i> Germar	Coleoptera, Hydraenidae	Hyd.gra
<i>Isoperla difformis</i> (Klapalek)	Plecoptera, Perlodidae	Iso.dif
<i>Leuctra</i> cf. <i>digitata</i> Kempny	Plecoptera, Leuctridae	Leu.dig
<i>Leuctra nigra</i> (Olivier)	Plecoptera, Leuctridae	Leu.nig
<i>Metacnephia bilineata</i> (Rubtsov)	Diptera, Simuliidae	Met.bil
<i>Micropsectra atrofasciata</i> -group	Diptera, Chironomidae	Mic.atr
<i>Orthocladius excavatus</i> -type	Diptera, Chironomidae	Ort.exc
<i>Orthocladius frigidus</i> (Zetterstedt)	Diptera, Chironomidae	Ort.fri
<i>Orthocladius olivaceus</i> -type	Diptera, Chironomidae	Ort.oli
<i>Orthocladius rhyacobius</i> -group	Diptera, Chironomidae	Ort.rhy
<i>Orthocladius rivicola</i> Kieffer	Diptera, Chironomidae	Ort.riv
<i>Orthocladius rivulorum</i> Kieffer	Diptera, Chironomidae	Ort.riu
<i>Paratrachocladius skirwithensis</i> (Edwards)	Diptera, Chironomidae	Par.ski
<i>Philopotamus montanus</i> (Donovan)	Trichoptera, Philopotamidae	Phi.mon
<i>Plectrocnemia conspersa</i> (Curtis)	Trichoptera, Polycentropodidae	Ple.con
<i>Potamophylax</i> cf. <i>cingulatus</i> (Stephens)	Trichoptera, Limnephilidae	Pot.cin
<i>Potthastia longimana</i> Kieffer	Diptera, Chironomidae	Pot.lon
<i>Prosimulium hirtipes</i> (Fries)	Diptera, Simuliidae	Pro.hir
<i>Rhyacophila nubila</i> Zetterstedt	Trichoptera, Rhyacophilidae	Rhy.nub
<i>Simulium monticola</i> Friederichs	Diptera, Simuliidae	Sim.mon
<i>Simulium murmanum</i> Enderlein	Diptera, Simuliidae	Sim.mur
<i>Siphonoperla burmeisteri</i> (Pictet)	Plecoptera, Chloroperlidae	Sip.bur
<i>Stegopterna trigonium</i> (Lundstrom)	Diptera, Simuliidae	Ste.tri
<i>Thienemanniella majuscula</i> -type	Diptera, Chironomidae	Thi.maj
<i>Trissopelopia longimana</i> (Staeger)	Diptera, Chironomidae	Tri.lon
<i>Tvetenia bavarica</i> (Goetghebuer)	Diptera, Chironomidae	Tve.bav
<i>Tvetenia calvescens</i> (Edwards)	Diptera, Chironomidae	Tve.cal
<i>Tvetenia discoloripes</i> (Goetghebuer & Thienemann)	Diptera, Chironomidae	Tve.dis

Table S2. Species traits considered and site occupancy (sites). Body size class (cm) refers to maximum size. Species ordered alphabetically by their codes (Table S1).

Species	Sites	Body size	Dispersal	Functional feeding group	Habit trait group
Ame.ino	33	1-2	low	Scraper	Swimmer
Amp.bor	23	0.5-1	low	Shredder	Sprawler
Amp.sul	28	0.5-1	low	Shredder	Sprawler
Apa.mul	6	1-2	low	Scraper	Sprawler
Bae.mut	38	1-2	low	Scraper	Swimmer
Bae.nig	24	1-2	low	Scraper	Swimmer
Bra.ris	24	1-2	low	Scraper	Sprawler
Car.cap	6	0.5-1	low	Predator	Clinger
Car.fus	11	0.5-1	low	Predator	Clinger
Con.bre	6	0-0.5	low	Gathering collector	Clinger
Cor.lob	12	0-0.5	low	Gathering collector	Sprawler
Diu.nan	10	2-4	low	Predator	Clinger
Elm.aen	14	0-0.5	low	Scraper	Clinger
Eph.aro	31	0.5-1	low	Gathering collector	Clinger
Euk.boe	8	0-0.5	low	Gathering collector	Sprawler
Euk.bre	8	0-0.5	low	Gathering collector	Sprawler
Euk.dev	33	0-0.5	low	Gathering collector	Sprawler
Hel.fer	41	0.5-1	high	Filtering collector	Clinger
Hep.dal	31	1-2	low	Scraper	Clinger
Hyd.gra	7	0-0.5	low	Scraper	Clinger
Iso.dif	42	1-2	low	Predator	Clinger
Leu.dig	31	1-2	low	Shredder	Sprawler
Leu.nig	20	1-2	low	Shredder	Sprawler
Met.bil	32	0.5-1	high	Filtering collector	Clinger
Mic.atr	43	0.5-1	low	Gathering collector	Sprawler
Ort.exc	7	0.5-1	low	Gathering collector	Sprawler
Ort.fri	26	0.5-1	low	Gathering collector	Sprawler
Ort.oli	12	0.5-1	low	Gathering collector	Sprawler
Ort.rhy	19	0.5-1	low	Gathering collector	Sprawler
Ort.riu	9	0.5-1	low	Gathering collector	Sprawler
Ort.riv	44	0.5-1	low	Gathering collector	Sprawler
Par.ski	9	0.5-1	low	Gathering collector	Sprawler
Phi.mon	18	1-2	low	Filtering collector	Clinger
Ple.con	14	2-4	high	Predator	Clinger
Pot.cin	9	2-4	high	Shredder	Clinger
Pot.lon	11	0.5-1	low	Gathering collector	Sprawler
Pro.hir	45	0.5-1	high	Filtering collector	Clinger
Rhy.nub	51	2-4	low	Predator	Clinger
Sim.mon	41	0.5-1	high	Filtering collector	Clinger
Sim.mur	29	0.5-1	high	Filtering collector	Clinger
Sip.bur	10	1-2	low	Predator	Clinger
Ste.tri	6	0.5-1	high	Filtering collector	Clinger
Thi.maj	9	0-0.5	low	Gathering collector	Sprawler
Tri.lon	14	0.5-1	low	Predator	Sprawler
Tve.bav	15	0.5-1	low	Gathering collector	Sprawler
Tve.cal	31	0.5-1	low	Gathering collector	Sprawler
Tve.dis	17	0.5-1	low	Gathering collector	Sprawler

Table S3. Results of binomial GLMs. Deviance explained (adjusted D² values) is decomposed into pure local environment (E) effects, pure climate (C) effects, pure spatial (S) effects, and combined (E+C+S) pure and joint effects. The effect (+/-) of environmental variables on species distributions is also shown. Environmental and spatial (V) variables are ordered as selected by each model (climate variables not shown). Species codes are as in Table S1.

Species	E	C	S	E+C+S	Variables selected in binomial models
Ame.ino	11.2	5.5	22.2	41.7	Sand (-), Water temperature (-); V4, V8, V1, V9
Amp.bor	21.6	13.1	5.1	37.1	Stream width (+), Water temperature (+); V6
Amp.sul	10.8	7.4	9.5	43.6	Velocity (+), Cobble (+), Bank slope (+); V1
Apa.mul	31.9	7.0	20.6	74.1	Pebble (+); V1, V8
Bae.mut	24.2	1.8	8.4	40.8	Stream width (+); V2, V4
Bae.nig	12.1	1.1	2.7	17.2	Stream width (+); V2
Bra.ris	22.9	17.0	31.5	70.4	Water temperature (+); V3, V4
Car.cap	9.6	1.0	28.6	42.4	Water temperature (+), Bank slope (+); V1, V5
Car.fus	5.5	1.0	3.1	12.5	Depth (+); V9
Con.bre	25.7	3.4	4.6	46.0	Manganese (+); V6
Cor.lob	9.5	12.6	17.2	45.0	Shading (+), Moss (+); V2, V12
Diu.nan	31.6	34.4	20.2	53.1	Pebble (+), Velocity (+); V4
Elm.aen	8.8	0.7	7.5	16.6	Moss (-); V3
Eph.aro	11.3	3.9	7.4	29.1	Stream width (-), Bank slope (-); V3, V4
Euk.boe	13.9	8.6	0.7	30.8	Sand (+); V6
Euk.bre	26.9	1.2	16.9	53.9	Moss (+), Shading (+); V2
Euk.dev	1.1	5.8	8.0	17.9	Depth (-); V8
Hel.fer	16.8	2.1	13.3	33.2	Cobble (+); V10, V3
Hep.dal	50.2	3.5	13.3	67.2	Stream width (+); V2
Hyd.gra	14.1	0.9	10.5	29.7	Manganese (+); V3
Iso.dif	19.0	1.8	19.1	28.6	Shading (+); V1, V4
Leu.dig	33.6	3.7	2.8	44.3	Water temperature (-), Shading (+), Cobble (+); V9
Leu.nig	17.0	3.7	10.4	29.4	Shading (+); V9
Met.bil	7.5	5.9	10.2	26.9	Water temperature (+); V13, V4
Mic.atr	14.2	1.3	8.8	33.9	Water temperature (-); V7, V1
Ort.exc	5.7	2.4	38.6	60.1	Water temperature (-); V4, V5
Ort.fri	3.0	0.4	2.7	19.4	Cobble (+), Pebble (+); V4
Ort.oli	13.4	0.4	6.9	20.7	Nitrogen (-); V3
Ort.rhy	6.8	12.4	21.7	44.5	Bank height (-), W. temp. (-), Shad. (+); V1, V3, V7
Ort.riu	2.5	0.4	6.9	10.5	Depth (+); V3
Ort.riv	9.1	28.9	23.5	55.5	Nitrogen (-); V10, V12
Par.ski	22.9	1.2	5.1	37.5	Moss (+), Bank slope (-); V7
Phi.mon	3.1	0.7	13.1	21.0	Shading (+); V1, V3
Ple.con	5.6	0.4	5.0	11.0	Shading (+); V1
Pot.cin	13.3	2.1	2.0	21.3	Stream width (-), Moss (+); V2
Pot.lon	6.0	1.1	5.2	15.3	Cobble (+); V3
Pro.hir	37.7	2.2	19.1	91.7	Shading (+), Nitrogen (-), Pebble (+); V9, V12
Rhy.nub	14.4	11.0	41.2	72.2	Pebble (+); V1
Sim.mon	8.8	1.6	9.2	30.2	Nitrogen (-); V3, V8
Sim.mur	12.0	1.8	8.5	22.2	Water temperature (+); V7
Sip.bur	66.7	20.1	1.6	82.1	Stream width (+); V3
Ste.tri	2.0	1.7	6.5	25.0	Water temperature (+); V3
Thi.maj	18.8	2.0	15.4	30.4	Moss (+); V10
Tri.lon	11.0	0.7	12.5	22.0	Manganese (-); V5
Tve.bav	20.8	6.2	2.2	28.1	Moss (+); V12
Tve.cal	16.8	3.3	15.3	33.2	Cobble (+), Manganese (+); V6, V3
Tve.dis	5.8	5.2	28.6	56.0	Shading (+), Cobble (+), Bank height (-); V1, V8
Mean	16.1	5.4	12.6	37.8	
Minimum	1.1	0.4	0.7	10.5	
Maximum	66.7	34.4	41.2	91.7	

Table S4. Taxonomic (TAX) and trait (TRA) vectors from Principal Coordinate Analysis (PCO), with corresponding scores for each species (codes of Table S1 in alphabetical order).

Species	TAX-1	TAX-2	TAX-3	TAX-4	TRA-1	TRA-2	TRA-3	TRA-4
Ame.ino	-0.000672	-0.001588	-0.000936	-0.002100	-0.0293	0.3933	0.1122	0.1866
Amp.bor	-0.000883	0.001971	0.000031	-0.000206	-0.1962	-0.0124	-0.2275	0.0800
Amp.sul	-0.000883	0.001971	0.000031	-0.000206	-0.1962	-0.0124	-0.2275	0.0800
Apa.mul	-0.000667	-0.001318	-0.000151	0.002842	-0.1949	0.2649	-0.0263	0.1285
Bae.mut	-0.000710	-0.001887	-0.001175	-0.002686	-0.0293	0.3933	0.1122	0.1866
Bae.nig	-0.000710	-0.001887	-0.001175	-0.002686	-0.0293	0.3933	0.1122	0.1866
Bra.ris	-0.000836	0.001659	0.000025	-0.000161	-0.1949	0.2649	-0.0263	0.1285
Car.cap	0.000627	0.000177	-0.000519	0.000067	0.1920	0.0249	0.0730	-0.2107
Car.fus	0.000627	0.000177	-0.000519	0.000067	0.1920	0.0249	0.0730	-0.2107
Con.bre	0.000615	0.000167	-0.000484	0.000062	0.0047	-0.1903	0.2542	-0.1198
Cor.lob	0.000615	0.000167	-0.000484	0.000062	-0.3372	-0.1323	0.0702	-0.0207
Diu.nan	-0.000864	0.001841	0.000028	-0.000186	0.2721	0.2469	-0.0171	-0.2712
Elm.aen	-0.000564	-0.000560	-0.000025	0.000208	0.1171	0.0981	0.3129	0.0977
Eph.aro	-0.000672	-0.001588	-0.000936	-0.002100	0.0407	-0.1447	0.1404	-0.0843
Euk.boe	0.000640	0.000188	-0.000560	0.000073	-0.3372	-0.1323	0.0702	-0.0207
Euk.bre	0.000640	0.000188	-0.000560	0.000073	-0.3372	-0.1323	0.0702	-0.0207
Euk.dev	0.000640	0.000188	-0.000560	0.000073	-0.3372	-0.1323	0.0702	-0.0207
Hel.fer	0.000222	-0.000436	0.002914	-0.000492	0.4740	-0.2543	-0.0050	0.1128
Hep.dal	-0.000672	-0.001588	-0.000936	-0.002100	0.1923	0.2710	0.1300	0.1053
Hyd.gra	-0.000564	-0.000560	-0.000025	0.000208	0.1171	0.0981	0.3129	0.0977
Iso.dif	-0.000864	0.001841	0.000028	-0.000186	0.2259	0.1929	0.0156	-0.1839
Leu.dig	-0.000883	0.001971	0.000031	-0.000206	-0.1853	0.1657	-0.3076	0.0424
Leu.nig	-0.000883	0.001971	0.000031	-0.000206	-0.1853	0.1657	-0.3076	0.0424
Met.bil	0.000222	-0.000436	0.002914	-0.000492	0.4740	-0.2543	-0.0050	0.1128
Mic.atr	0.000615	0.000167	-0.000484	0.000062	-0.2591	-0.0955	-0.0218	-0.0117
Ort.exc	0.000681	0.000230	-0.000732	0.000097	-0.2591	-0.0955	-0.0218	-0.0117
Ort.fri	0.000681	0.000230	-0.000732	0.000097	-0.2591	-0.0955	-0.0218	-0.0117
Ort.oli	0.000681	0.000230	-0.000732	0.000097	-0.2591	-0.0955	-0.0218	-0.0117
Ort.rhy	0.000681	0.000230	-0.000732	0.000097	-0.2591	-0.0955	-0.0218	-0.0117
Ort.riu	0.000681	0.000230	-0.000732	0.000097	-0.2591	-0.0955	-0.0218	-0.0117
Ort.riv	0.000681	0.000230	-0.000732	0.000097	-0.2591	-0.0955	-0.0218	-0.0117
Par.ski	0.000615	0.000167	-0.000484	0.000062	-0.2591	-0.0955	-0.0218	-0.0117
Phi.mon	-0.000667	-0.001318	-0.000151	0.002842	0.2777	0.0496	0.0362	0.0862
Ple.con	-0.000667	-0.001318	-0.000151	0.002842	0.6379	0.0928	-0.1271	-0.2234
Pot.cin	-0.000667	-0.001318	-0.000151	0.002842	0.5730	0.0584	-0.3733	0.1310
Pot.lon	0.000615	0.000167	-0.000484	0.000062	-0.2591	-0.0955	-0.0218	-0.0117
Pro.hir	0.000222	-0.000436	0.002914	-0.000492	0.4740	-0.2543	-0.0050	0.1128
Rhy.nub	-0.000667	-0.001318	-0.000151	0.002842	0.2721	0.2469	-0.0171	-0.2712
Sim.mon	0.000226	-0.000460	0.003126	-0.000531	0.4740	-0.2543	-0.0050	0.1128
Sim.mur	0.000226	-0.000460	0.003126	-0.000531	0.4740	-0.2543	-0.0050	0.1128
Sip.bur	-0.000836	0.001659	0.000025	-0.000161	0.2259	0.1929	0.0156	-0.1839
Ste.tri	0.000222	-0.000436	0.002914	-0.000492	0.4740	-0.2543	-0.0050	0.1128
Thi.maj	0.000615	0.000167	-0.000484	0.000062	-0.3372	-0.1323	0.0702	-0.0207
Tri.lon	0.000615	0.000167	-0.000484	0.000062	-0.1479	0.0545	-0.1030	-0.2533
Tve.bav	0.000640	0.000188	-0.000560	0.000073	-0.2591	-0.0955	-0.0218	-0.0117
Tve.cal	0.000640	0.000188	-0.000560	0.000073	-0.2591	-0.0955	-0.0218	-0.0117
Tve.dis	0.000640	0.000188	-0.000560	0.000073	-0.2591	-0.0955	-0.0218	-0.0117

Table S5. Statistical significance (P -values) of the correlation between site occupancy, species traits, trait vectors, and taxonomic vectors.

Depending on the continuous or categorical nature of the variables involved, P -values refer to either the significance of the Spearman correlation coefficient, to Kruskal-Wallis tests, to Mann-Whitney tests, or to Fisher's exact tests on a contingency table. Although P -values from these various analyses are not strictly comparable, they provide an overall idea of the degree to which two variables are correlated. Nsites, site occupancy; BS, body size; DP, dispersal potential; FFG, functional feeding group; HTG, habit trait group; TRA, trait vector; TAX, taxonomic vector. Vectors are multivariate axes of a Principal Coordinate Analysis combining either trait or taxonomic variables. Significant P -values ($P < 0.05$) are shown in boldface; marginally significant ($P < 0.10$) in italics.

	Nsites	BS	DP	FFG	HTG	TAX-1	TAX-2	TAX-3	TAX-4	TRA-1	TRA-2	TRA-3	TRA-4
Nsites	-												
BS	0.108	-											
DP	0.275	0.043	-										
FFG	0.468	<0.001	<0.001	-									
HTG	0.275	0.024	0.001	<0.001	-								
TAX-1	0.112	<0.001	0.599	<0.001	0.009	-							
TAX-2	0.547	0.492	0.019	<0.001	<0.001	0.376	-						
TAX-3	0.637	0.750	<0.001	<0.001	<0.001	<0.001	0.373	-					
TAX-4	0.007	0.022	0.151	<i>0.072</i>	0.008	<0.001	0.496	0.223	-				
TRA-1	0.224	0.004	<0.001	<0.001	<0.001	<0.001	0.002	<0.001	0.083	-			
TRA-2	0.808	<0.001	0.003	<0.001	0.015	<0.001	0.793	0.220	0.591	0.337	-		
TRA-3	0.784	0.002	0.467	0.002	<0.001	0.869	<0.001	<i>0.085</i>	0.103	0.382	0.626	-	
TRA-4	0.101	0.028	0.014	<0.001	0.015	0.038	0.011	0.250	0.019	0.120	0.586	0.783	-

Fig. S1. A map of the study area located in the Tenokoki drainage basin. Both the location of the Tenojoki basin at continental scale (A), and the details of the river network (B), are shown. All sampling points are located in tributary streams of the River Tenojoki or the River Utsjoki. Dashed lines show the border between Norway and Finland. Note that only parts of the stream networks on the Norwegian side of the border are shown for clarity.

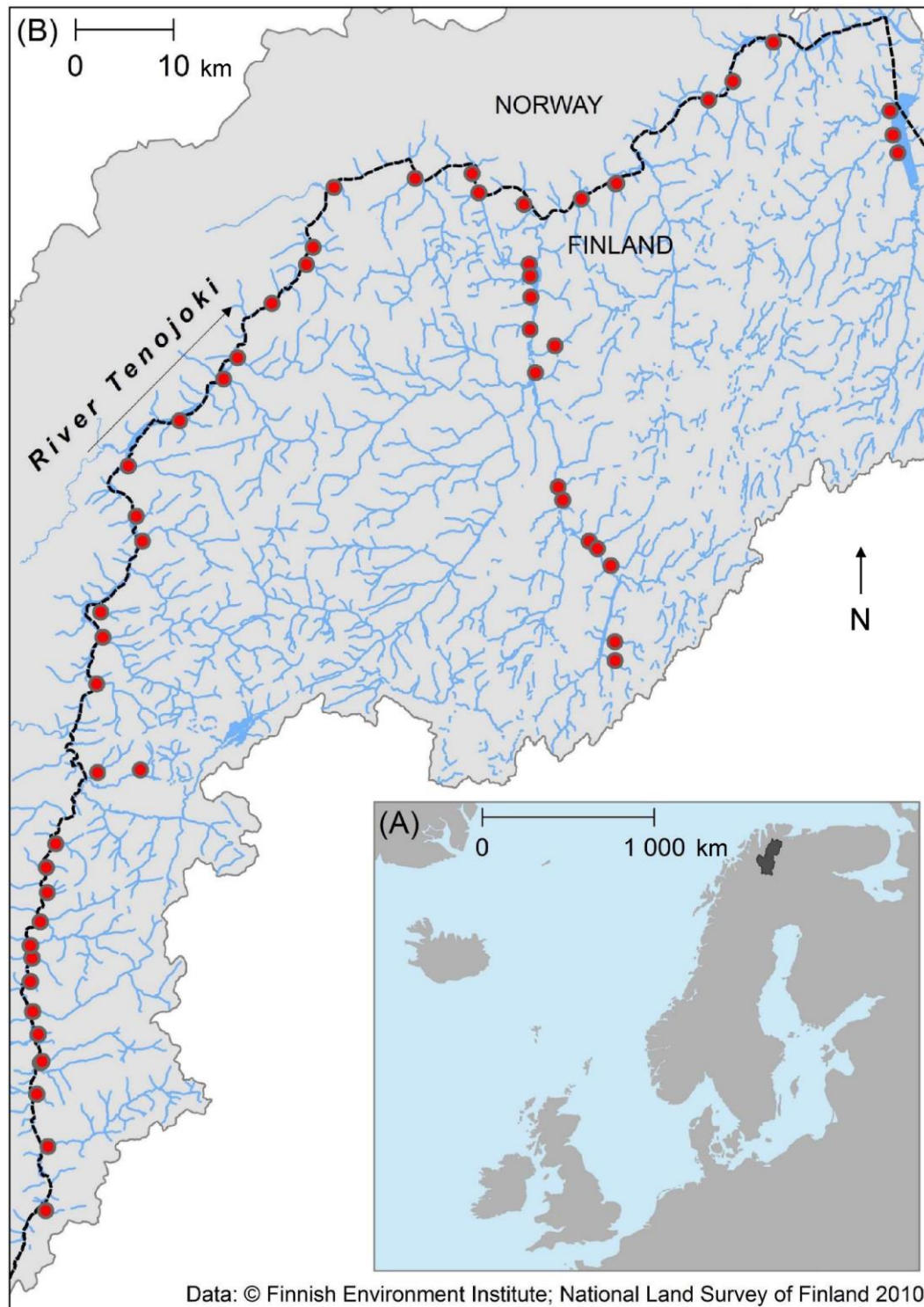


Fig. S2. Eigenvalues from taxonomic Principal Coordinate Analysis (PCO) based on taxonomic distances between species. The first four eigenvectors were selected in the comparative analysis across species.

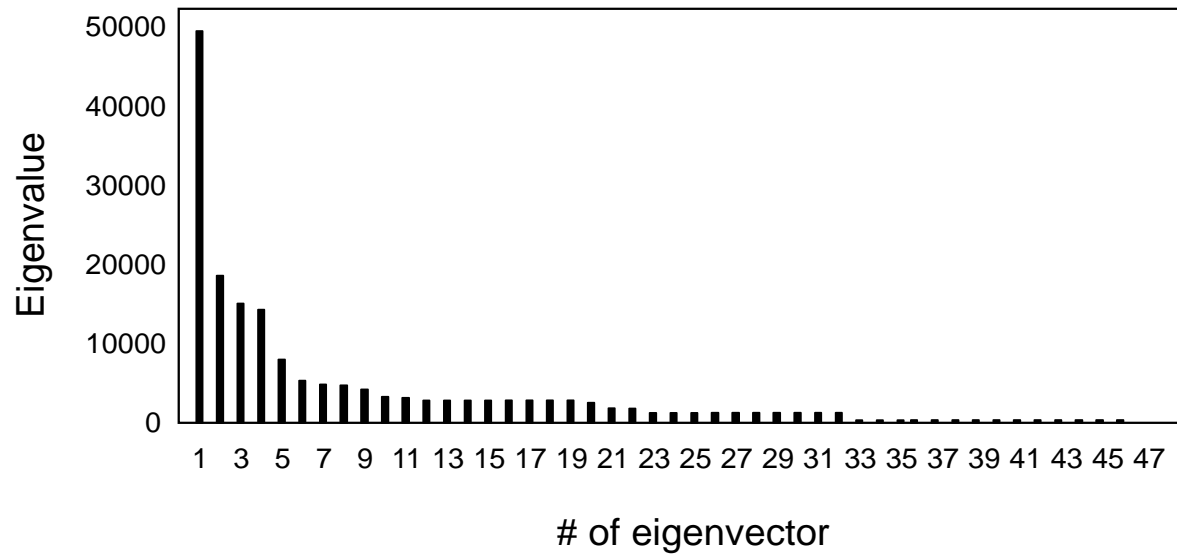


Fig. S3 (see next page). Results of binomial GLMs (adjusted D^2 values) in relation to site occupancy and species traits. From left to right, columns refer to site occupancy, body size classes (cm), dispersal potential, functional feeding groups, and habit trait groups, respectively. The adjusted D^2 values of binomial GLMs, result from pure local environmental (E), pure climate (C), or pure spatial (S) effects, as well as from a combination of the pure and joint effects of the three subsets of variables (E+C+S). Numbers inside graphs refer to P -values of univariate effects of variables on the deviance explained: beta regressions for site occupancy; Kruskal-Wallis tests for body size classes, functional feeding groups, and habit trait groups; and Mann-Whitney tests for dispersal potential. Significance ($P < 0.05$) is indicated with an asterisk. Letters inside graphs indicate significant differences ($P < 0.05$, Mann-Whitney tests) for pairwise comparisons, in case of statistical significance of the Kruskal-Wallis test. Abbreviations: Filt, filtering collector; Gath, gathering collector, Pred, predator; Scrap, scraper; Shred, shredder; Clin, clinger; Spra, sprawler; Swim, swimmer.

Fig. S3

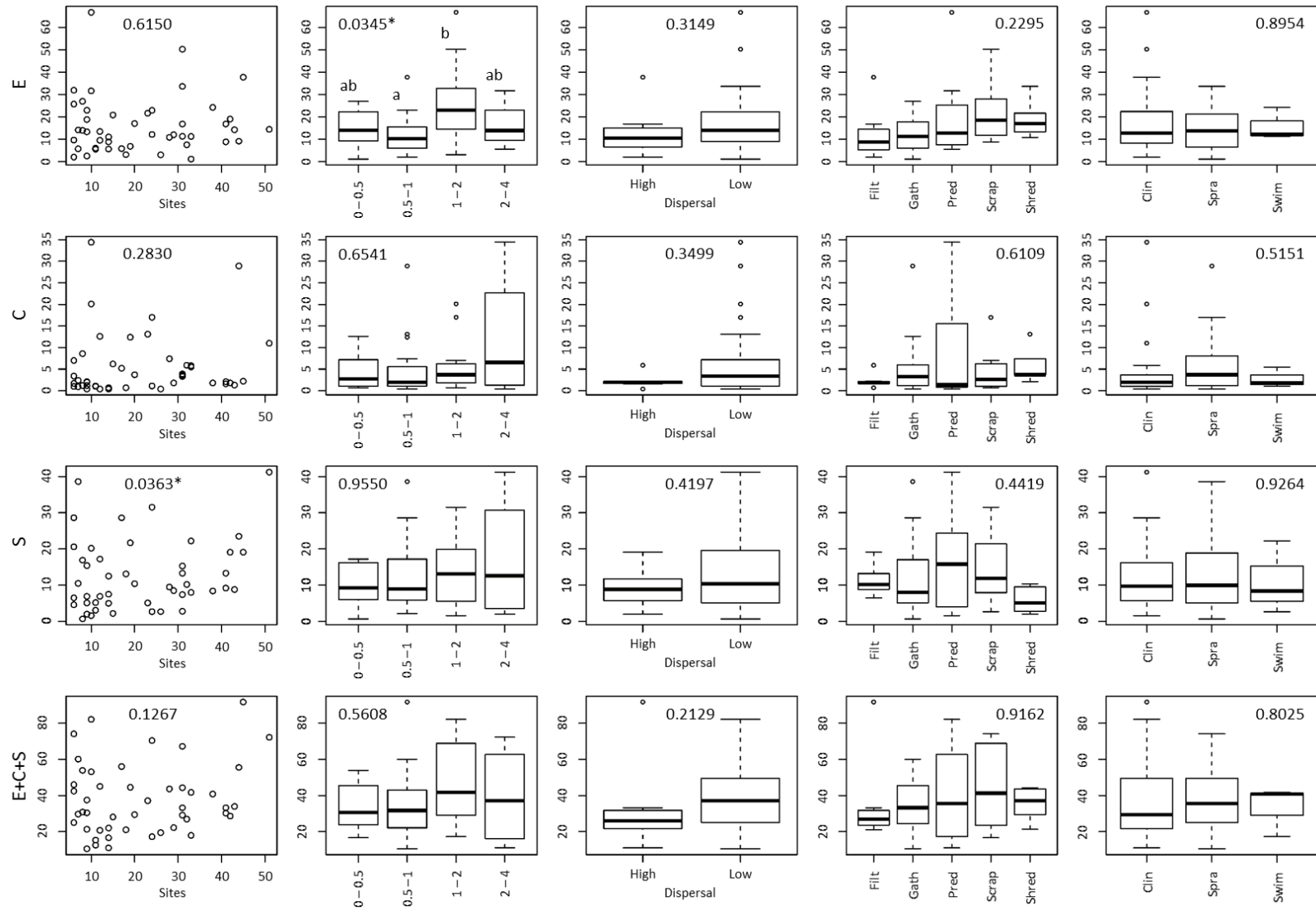


Fig. S4. Comparison of site occupancy values between different insect groups. a) broad taxonomic groups; b) functional feeding groups; c) habit trait groups. Numbers inside boxes indicate the *P*-value of the Kruskal-Wallis test of each analysis, with the asterisk indicating marginal statistical significance ($P < 0.10$). In case the Kruskal-Wallis test was (marginally) significant, letters inside boxes indicate marginal statistical differences ($P < 0.10$; Mann-Whitney test; in a) the Simuliidae differed from the Chironomidae but not from the other groups).

